

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-108681

(43) 公開日 平成10年(1998) 4月28日

(51) IntCl. ⁹	識別記号	F I	
C 1 2 N 15/09	Z N A	C 1 2 N 15/00	Z N A A
A 6 1 K 38/46		A 6 1 K 48/00	
48/00		C 0 7 H 21/04	B
C 0 7 H 21/04		C 1 2 N 9/64	Z
C 1 2 N 5/10		C 1 2 Q 1/68	A

審査請求 未請求 請求項の数47 F D 外国語出願 (全 189 頁) 最終頁に続く

(21) 出願番号 特願平9-170867

(22) 出願日 平成9年(1997) 5月23日

(31) 優先権主張番号 60/019942

(32) 優先日 1996年6月14日

(33) 優先権主張国 米国 (US)

(31) 優先権主張番号 60/020273

(32) 優先日 1996年6月17日

(33) 優先権主張国 米国 (US)

(31) 優先権主張番号 60/026083

(32) 優先日 1996年8月28日

(33) 優先権主張国 米国 (US)

(71) 出願人 591002957
スミスクライン・ビーチャム・コーポレイ
ション
SMITHKLINE BEECHAM
CORPORATION
アメリカ合衆国ペンシルベニア州19406-
0939、キング・オブ・ブルシア、スウェー
ドランド・ロード709番

(72) 発明者 マーク・ディ・アダムス
アメリカ合衆国20997メリーランド州ノー
ス・ポトマック、ドゥフィーフ・ドライブ
15205番

(74) 代理人 弁理士 青山 稔 (外1名)
最終頁に続く

(54) 【発明の名称】 カテプシンK遺伝子

(57) 【要約】

【課題】 結合組織の生理学的および病理学的分解にて機能を発揮するプロテアーゼである、カテプシンについて新しく産生等することが望まれている。

【解決手段】 本発明はカテプシンKポリペプチド、そのポリペプチドをコードするポリヌクレオチド、特に該ポリヌクレオチドを発現することによるポリペプチドの産生方法、並びにポリペプチドのアゴニストおよびアンタゴニストを提供するものである。本発明はさらに、研究、診断および臨床分野に関連する、このようなポリヌクレオチド、ポリペプチド、アゴニストおよびアンタゴニストの利用方法をも提供する。

AL4

【特許請求の範囲】

【請求項1】 配列番号：1の配列と少なくとも80%同一である配列；

配列番号：1の配列と少なくとも85%同一である配列；

配列番号：1の配列と少なくとも90%同一である配列；

配列番号：1の配列と少なくとも95%同一である配列；および

配列番号：1の配列と少なくとも97%同一である配列；

からなる群より選択される領域を有してなる単離ポリヌクレオチド。

【請求項2】 領域がゲノムDNAまたはcDNAである請求項1記載の単離ポリヌクレオチド。

【請求項3】 カテプシンKエンハンサーまたはプロモーターを有してなる単離ポリヌクレオチド。

【請求項4】 配列番号：1の配列を有する請求項3記載の単離ポリヌクレオチド。

【請求項5】 カテプシンKボリアデニル化領域を有してなる単離ポリヌクレオチド。

【請求項6】 配列番号：1の配列を有する請求項5記載の単離ポリヌクレオチド。

【請求項7】 カテプシンKイントロンを有してなる単離ポリヌクレオチド。

【請求項8】 配列番号：1の配列を有する請求項7記載の単離ポリヌクレオチド。

【請求項9】 イントロン1、2、3、4、5、6および7からなる群より選択される配列を有してなる単離ポリヌクレオチド。

【請求項10】 イントロン1、2、3、4、5、6および7からなる群より選択される配列を有してなるポリヌクレオチドによりコードされる単離ポリペプチド。

【請求項11】 カテプシンKエキソンを有してなる単離ポリヌクレオチド。

【請求項12】 配列番号：1の配列を有する請求項11記載の単離ポリヌクレオチド。

【請求項13】 エキソン1、2、3、4、5、6、7および8からなる群より選択される配列を有してなる単離ポリヌクレオチド。

【請求項14】 エキソン1、2、3、4、5、6、7および8からなる群より選択される配列を有してなるポリヌクレオチドによりコードされる単離ポリペプチド。

【請求項15】 1-3、1-4、1-5、1-6、1-7、1-8、2-4、2-5、2-6、2-7、2-8、3-4、3-5、3-6、3-7、3-8、4-5、4-6、4-7、4-8、5-7、5-8および6-8からなる群より選択されるエキソン-エキソン対を有してなる単離ポリヌクレオチド。

【請求項16】 1-3、1-4、1-5、1-6、1-

7、1-8、2-4、2-5、2-6、2-7、2-8、3-4、3-5、3-6、3-7、3-8、4-5、4-6、4-7、4-8、5-7、5-8および6-8からなる群より選択されるエキソン-エキソン対を有してなるポリヌクレオチドによりコードされる単離ポリペプチド。

【請求項17】 ATCC寄託番号98035のヒトcDNAと少なくとも80%同一である配列；ATCC寄託番号98035のヒトcDNAと少なくとも85%同一である配列；ATCC寄託番号98035のヒトcDNAと少なくとも90%同一である配列；ATCC寄託番号98035のヒトcDNAと少なくとも95%同一である配列；およびATCC寄託番号98035のヒトcDNAと少なくとも97%同一である配列；からなる群より選択される領域を有してなる単離ポリヌクレオチド。

【請求項18】 (a) 配列番号：20に示すアミノ酸配列を有してなるポリペプチドをコードするポリヌクレオチドと少なくとも70%の同一性を有するポリヌクレオチド；(b) 配列番号：20に示すアミノ酸1からアミノ酸329を有してなるポリペプチドをコードするポリヌクレオチドと少なくとも70%の同一性を有するポリヌクレオチド；(c) (a)または(b)のポリヌクレオチドにハイブリダイズし、そのポリヌクレオチドと少なくとも70%相補的であるポリヌクレオチド；および(d) (a)、(b)または(c)のポリヌクレオチドのポリヌクレオチド断片が配列番号：1の少なくとも30の連続した塩基を有してなる、(a)、(b)または(c)のポリヌクレオチドのポリヌクレオチド断片；からなる群より選択される構成因子を有してなる単離ポリヌクレオチド。

【請求項19】 領域が、ATCC寄託番号98035のヒトcDNA挿入物におけるカテプシンKをコードする領域である、請求項1記載の単離ポリヌクレオチド。

【請求項20】 宿主細胞にて機能的に連結した請求項1に記載のポリヌクレオチドの発現に有効なシス作用制御要素を有してなる発現ベクター。

【請求項21】 制御要素が宿主細胞におけるポリヌクレオチドの発現を誘発するのに効果的である請求項20記載発現ベクター。

【請求項22】 宿主細胞にて機能的に連結した請求項18に記載のポリヌクレオチドの発現に有効なシス作用制御要素を有してなる発現ベクター。

【請求項23】 その中に請求項22に記載の発現ベクターを発現するように組み込まれている宿主細胞。

【請求項24】 その中に請求項18に記載の発現ベクターを発現するように組み込まれている宿主細胞。

【請求項25】 宿主細胞にて請求項1に記載のポリヌクレオチドを発現する工程を有してなるポリペプチドの製造方法。

【請求項26】 宿主細胞にて請求項18に記載のポリヌクレオチドを発現する工程を有してなるポリペプチドの製造方法。

【請求項27】 配列番号：20のアミノ酸配列の連続領域と配列が同一である15またはそれ以上のアミノ酸のポリペプチド。

【請求項28】 ATCC寄託番号98035のヒトcDNAにコードされるポリペプチドのアミノ酸配列の連続領域と配列が同一である15またはそれ以上のアミノ酸のポリペプチド。

【請求項29】 配列番号：20のアミノ酸配列の連続領域と配列が同一である50またはそれ以上のアミノ酸のポリペプチド。

【請求項30】 配列番号：1のポリペプチドと配列が少なくとも90%同一であるポリヌクレオチドの連続領域と配列が同一である25またはそれ以上のヌクレオチドのポリヌクレオチド。

【請求項31】 50またはそれ以上のヌクレオチドの請求項23記載のポリヌクレオチド。

【請求項32】 75またはそれ以上のヌクレオチドの請求項31記載のポリヌクレオチド。

【請求項33】 請求項20に記載のベクターで遺伝子操作した宿主細胞。

【請求項34】 該DNAによりコードされるポリペプチドを請求項33記載の宿主細胞から発現することからなるポリペプチドの産生方法。

【請求項35】 請求項20に記載のベクターで細胞を遺伝子操作することからなるポリペプチドの発現能を有する細胞の産生方法。

【請求項36】 試料中のカテプシンKをコードするポリヌクレオチドの測定法であって、カテプシンKをコードするポリヌクレオチドに特異的なプローブが該ポリヌクレオチドと特異的にハイブリダイズするのに有効な条件下で、該プローブを試料とハイブリッド形成させ、該試料中のポリヌクレオチドに対する該プローブのハイブリダイゼーションを測定する工程からなり、ここに、該プローブは配列番号：1のポリヌクレオチド配列と少なくとも90%同一の20またはそれ以上の塩基対の領域の配列を有することからなる方法。

【請求項37】 試料中のカテプシンKをコードするポリヌクレオチドの測定方法であって、カテプシンKをコードするポリヌクレオチドに特異的なプローブが該ポリヌクレオチドと特異的にハイブリダイズするのに有効な条件下で、該プローブを試料とハイブリッド形成させ、該試料中のポリヌクレオチドに対する該プローブのハイブリダイゼーションを測定する工程からなり、ここに、該プローブはATCC寄託番号98035のヒトcDNA挿入物のポリヌクレオチド配列と少なくとも

90%同一の20またはそれ以上の塩基対の領域の配列を有することからなる方法。

【請求項38】 試料中、配列番号：1の1から135の残基のアミノ酸配列と配列が少なくとも90%同一の領域を有してなるポリペプチドを検出する方法であって、

特異的結合に効果的な条件下で該ポリペプチドに特異的に結合する試薬を試料と一緒にインキュベーションし；および該試料中のポリペプチドへの該試薬の結合を測定することからなる検定方法。

【請求項39】 カテプシンKのポリヌクレオチドの異常発現により特徴づけられる疾患の診断方法であって、特異的ハイブリダイゼーションに効果的な条件下で、配列番号：1の1から329のアミノ酸をコードするRNAまたはDNAに配列が少なくとも90%同一である領域を有してなるポリヌクレオチドに対して特異的なプローブをハイブリッド形成し；該試料中の該ポリヌクレオチドへの該プローブのハイブリダイゼーションを測定することからなる方法。

【請求項40】 カテプシンKのポリペプチドの異常発現により特徴づけられる疾患の診断方法であって、特異的結合に効果的な条件下で配列番号：1の1から329の残基のアミノ酸配列と配列が少なくとも90%同一である領域を有してなるポリペプチドに特異的に結合する試薬を試料と一緒にインキュベーションし；および該試料中の該ポリペプチドへの該試薬の結合を測定することからなる方法。

【請求項41】 請求項13に記載のポリペプチドの活性化を阻害する化合物。

【請求項42】 請求項13に記載の治療上有効量のポリペプチドを患者に投与することからなるカテプシンKを必要とする患者の治療方法。

【請求項43】 治療上有効量のポリペプチドを、該ポリペプチドをコードするDNAを患者に付与し、in vivoにて該ポリペプチドを発現することにより投与する請求項42記載の方法。

【請求項44】 カテプシンKポリペプチドを阻害する必要がある患者の治療方法であって、治療上有効量の請求項41記載の化合物を該患者に投与することからなる方法。

【請求項45】 請求項13に記載のポリペプチドの発現不足に關与する疾患またはその疾患に対する感受性の診断方法であって、該ポリペプチドをコードする核酸配列における変異を測定することからなる方法。

【請求項46】 宿主由来の試料中の請求項13に記載のポリペプチドの存在について分析することからなる診断方法。

【請求項47】 請求項13に記載のポリペプチドに結合し、該ポリペプチドの活性化を阻害する化合物の同定

方法であって、

該ポリペプチドに対する受容体（化合物の受容体への結合にตอบสนองして、検出可能な信号を付与しうる第二の成分と結合する）をその表面で発現する細胞を、分析的に検出可能なカテプシンKのポリペプチドおよび化合物と、受容体への結合を可能とする条件下で接触させ、カテプシンKと受容体との相互作用から生じる信号の不在を検出することにより、化合物が受容体に結合し、その受容体を阻害するかどうかを測定することからなる同定方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は一つの態様として、新規に同定されたポリヌクレオチドおよびポリペプチド；ポリヌクレオチドおよびポリペプチドの変異型および誘導体；ポリヌクレオチドおよびポリペプチド、およびそれらの変異型および誘導体の製造方法；ポリペプチドのアゴニストおよびアンタゴニスト；並びにポリヌクレオチド、ポリペプチド、変異型、誘導体、アゴニストおよびアンタゴニストの使用に関するものである。とりわけ、これらのおよびその他の点に関して、本発明はヒトカテプシンKのポリヌクレオチドおよびポリヌクレオチド、特にカテプシンKのゲノム配列、とりわけプロモータおよびイントロン配列に関する。

【0002】

【従来の技術および発明が解決しようとする課題】 骨の吸収は細胞外マトリックスの無機および有機成分の両方が同時に除去されて起こる。これは主に、破骨細胞の波状の辺縁で覆われた、酸性の食胞融解小体様細胞外部分で起こる。Barronら、*J. Cell Biol.*, 101:2210-2222 (1985)。破骨細胞は骨吸収において主要な役割を果たす多核巨細胞である。破骨細胞は、骨表面に付着して、破骨細胞と骨マトリックスとの間で酸性の微小環境を創り出す。この酸性の微小環境は骨の無機および有機成分を可溶化する。有機成分主にI型コラーゲンはプロテアーゼ消化により可溶化されると考えられている。システインプロテイナーゼは骨の有機成分の分解において重要な役割を果たしうるといふ証拠がある。システインプロテイナーゼの中では、カテプシンB、L、HおよびSが酸性条件下でI型コラーゲンを分解できる。Etherington, D. J., *Biochem. J.*, 127, 685-692 (1972)。カテプシンLは、アソカゼイン、エラスチンおよびコラーゲンの加水分解能に関して、最も活性なライソソームシステインプロテアーゼである。

【0003】 カテプシンは、結合組織において、正常な生理学的および病理学的分解に機能するプロテアーゼである。カテプシンは細胞内のタンパク質分解およびターンオーバー、骨の再形成、およびプロホルモンの活性化において主要な役割を果たす。Marix, J. L., *Science*, 235:285-286 (1987)。カテプシンB、H、LおよびS

は、パパインのスーパーファミリーである偏在性の発現をするライソソームシステインプロテイナーゼである。これらはヒトの腎臓、肝臓、肺および脾臓などの多くの組織に構成成分のレベルで見出される。カテプシンは、糸球体腎炎、関節炎および癌転移の誘起に病理学的な役割を果たすこともある。Sloan, B. F. および Honn, K. V., *Cancer Metastasis Rev.*, 3:249-263 (1984)。腫瘍細胞では、カテプシンLおよびB mRNA並びにタンパク質のレベルが非常に上昇しているのが認められる。カテプシンL mRNAはまた、腫瘍促進物質および成長因子で処理した繊維芽細胞においても誘起される。Kane, S. E. および Gottesman, M. M., *Cancer Biology*, 1:127-136 (1990)。

【0004】 アルツハイマー病の脳における非システインプロテアーゼ、カテプシンDの遺伝子発現および細胞成分は、エンドソームのライソソーム系に初期上向制御（アップレギュレーション）するのが明白である。Cataldo, A. M. ら、*Neuron*, 14(3), 671-680 (1995)。骨吸収に関するインビトロ研究において、カテプシンLおよびBがこの組織の再形成を引き起こし得ることが示されている。これらのライソソームシステインプロテアーゼは、酸性条件下で、エラスチン、ラミニンおよびI型コラーゲンなどの細胞外マトリックスタンパク質を消化する。骨芽細胞により成就される骨再発生に先立ち、有機マトリックスを分解するために、破骨細胞にはこの作用が必要である。システインプロテイナーゼの天然および合成阻害物質には、このマトリックスの分解の阻害に有効であるものもある。カテプシンの単離およびそれらの骨吸収における役割が、熱心な研究対象となっている。最近、OC-2がウサギの骨の純粋な破骨細胞から単離された。OC-2は、構造的にカテプシンLおよびSに関連する、システインプロテイナーゼと成り得るものをコードすることが解っている。Tezuka, K. ら、*J. Biol. Chem.*, 269:1106-1109 (1994)。

【0005】 システインプロテイナーゼおよびコラーゲナーゼの阻害物質、Z-Phe-Ala-CHN₂の、単離破骨細胞の吸収活性に及ぼす影響が研究され、ぞうげ質における吸収ピットを阻害することが見出されている。Delaisse, J. M. ら、*Bone*, 8:305-313 (1987)。またインビトロでの骨吸収に及ぼすヒト組換えシステインC、システインプロテイナーゼ阻害物質の影響も評価されており、上皮小体ホルモンにより刺激された骨再吸収を著明に阻害することが示されている。Lerner, U. H. および Grubb Anders, *Journal of Bone and Mineral Research*, 7:433-439 (1989)。さらに、ヒトシステインプロテアーゼカテプシンLをコードするcDNAクローンは、大腸菌 (*E. coli*) においてT7発現系で高レベルで組換えにより製造および発現されている。組換えヒトプロカテプシンLは、うまく高レベルで発現され、プロカテプシンLおよび活性なプロセッシングしたカテプシン

L体の両方の形で精製された。カテプシンLの折り畳みおよび/またはプロセッシングにおけるペプチドの可能な機能に関する情報、およびプロテアーゼ活性のための酵素の軽鎖の必要性に関する情報が、これらの領域に構造的な改変を担う突然変異酵素の発現および精製により得られた。Smith, S. M. および Gottesman, M. M., J. Bio Chem., 264:20487-20495 (1989)。これらは機能的なヒトカテプシンSのビール酵母菌 (*Saccharomyces cerevisiae*) における発現および組換え酵素の特性化についても報告している。Bromme, D. ら, J. Biol. Chem., 268:4832-4838 (1993)。

【0006】

【課題を解決するための手段】上記およびその他の課題に対し、本発明の目的は、ポリペプチドとりわけ図26および27に示すアミノ酸配列とその他のタンパク質の既知アミノ酸配列、例えばウサギOC-2とヒトカテプシンOcDNAとの相同性により新規カテプシンKとして同定したものを提供することにある。Tezuka, K. ら, J. Biol. Chem., 269:1106-1109 (1994)。さらに本発明の目的は、カテプシンK、特に本明細書においてカテプシンKと称するポリペプチドをコードするポリヌクレオチドを提供することにある。

【0007】本発明のとりわけ好ましい態様において、ポリヌクレオチドは、図1-9【配列番号：1】またはATCC寄託番号98035（本明細書では寄託クローンと称する）のゲノムDNA（本明細書では「gDNA」と示す）に示す配列におけるヒトカテプシンKをコードする領域を包含する。本発明の態様に準じて、ヒトカテプシンKを、本発明のさらなる態様において、生物学的、診断学的、臨床的または治療的に有用なそれらの変異型、類似体もしくは誘導体、または変異型、類似体または誘導体の断片等のそれらの断片をコードする単離核酸分子、例えばmRNA類、cDNA類、ゲノムDNA類等を提供する。本発明のとりわけ好ましい態様は、ヒトカテプシンKの天然発生のアレル変異型である。

【0008】本発明の更なる目的は、カテプシンKポリペプチド、とりわけ例えば骨粗鬆症、パジェット病、ゴーチャーズ病、CNS炎症、アルツハイマー病、上皮小体亢進症、骨分解、転移性腫瘍、リウマチ性関節症、変形性関節症、歯周病および骨インプラントおよび骨プロシーゼス (protheses) とりわけ歯科インプラントの分解などを引き起こすまたはこれらに関与するヒトカテプシンKポリペプチドを提供することにある。本発明の態様に準じ、本明細書にてカテプシンKおよび生物学的、診断学的、または治療的に有用な断片、それらの変異型および誘導体、その断片の変異型および誘導体並びに前掲の類似体と称される、ヒト由来の新規ポリペプチドを提供する。

【0009】本発明のとりわけ好ましい態様は、ヒトカテプシンK遺伝子の天然発生のアレルによりコードされ

るヒトカテプシンKの変異型である。本発明のもう一つの目的は、前述のポリペプチド、ポリペプチド断片、変異型および誘導体、変異型および誘導体の断片、並びに前掲の類似体の製造方法を提供することにある。本発明のとりわけ好ましい態様は、外因性のヒトカテプシンKをコードするポリヌクレオチドを、発現可能な様に挿入した宿主細胞を宿主においてヒトカテプシンKを発現するような条件下で培養し、発現したポリペプチドを回収することを含む、前掲のカテプシンKポリペプチドの製造方法を提供する。

【0010】さらに本発明のもう一つの目的は、カテプシンK遺伝子を欠損する、または欠損が疑われる個体の薬物の反応性を測定する方法を提供することである。さらに本発明のもう一つの目的は、とりわけ生物学的、臨床的および治療的目的での研究のために前掲のポリペプチドおよびポリヌクレオチドを利用する産生物、組成物、過程および方法を提供することである。本発明の特定の好ましい態様において、産生物、組成物および方法、なかでも：細胞をカテプシンKポリペプチド、ポリヌクレオチドまたは本明細書に開示する抗体に曝露することによりインビトロ、エクソビオまたはインビボでカテプシンKをコードするmRNAまたはhnRNAからのカテプシンKポリペプチドを測定することによりカテプシンKの細胞中の発現を評価する方法；カテプシンKポリヌクレオチドにおける遺伝子変異および異常、例えば遺伝子および遺伝子制御配列の欠損を検定する方法；およびカテプシンK機能を増強するために、またはカテプシンK機能障害を治療するために、生体にカテプシンKポリペプチドまたはポリヌクレオチドを投与する方法を提供する。

【0011】本発明の特定の好ましい態様において、ヒトカテプシンK配列に特異的にハイブリダイズするプローブを提供する。本発明のさらに好ましい態様において、カテプシンKポリペプチドに対する抗体を提供する。この点に関して特に好ましい態様では、その抗体はヒトカテプシンKに対して高選択的である。本発明の他の態様では、カテプシンKアゴニストを提供する。とりわけ好ましいアゴニストは、カテプシンK結合分子、または受容体分子に結合する、およびカテプシンK誘起反応を引き出すまたは増強するカテプシンK擬似分子である。また、とりわけ好ましいアゴニストは、カテプシンKもしくはカテプシンKポリペプチド、またはその他のカテプシンK活性制御物質と相互作用し、それによりカテプシンKの効果またはカテプシンKの効果以上の効果を可能にするかまたは増強する分子である。

【0012】本発明のさらに別の態様において、カテプシンKアンタゴニストを提供する。とりわけ好ましいアンタゴニストは、カテプシンK受容体または結合分子に結合するが、一つまたは一つ以上のカテプシンK誘起反応を引き出さないようなカテプシンK擬似分子である。

また、とりわけ好ましいアンタゴニストは、カテプシンKの一つまたは一つ以上の効果を阻害するするために、カテプシンKと結合または相互作用する。アゴニストおよびアンタゴニストは、カテプシンKポリペプチドの作用を擬似、増強または阻害するために使用できる。これらは、例えば、骨粗鬆症、パジェット病、ゴーチャーズ病、CNS炎症、アルツハイマー病、上皮小体亢進症、骨分解、転移性腫瘍、および骨インプラントおよび骨プロシーズ (protheses) とりわけ歯科インプラントの分解の処置に用いることができる。このようなアンタゴニストは、とりわけ骨粗鬆症、パジェット病、ゴーチャーズ病、アルツハイマー病、上皮小体亢進症、骨分解、転移性腫瘍、CNS炎症、リウマチ性関節症、変形性関節症、歯周病および骨インプラントおよび骨プロシーズ (protheses) とりわけ歯科インプラントの分解の処置に有用である。

【0013】本発明の更なる態様は、インビトロで細胞に、エクソビポで細胞におよびインビポで細胞に、または多細胞生体に投与するための、カテプシンKポリヌクレオチドまたはカテプシンKポリペプチドを含む組成物を提供する。本発明の特に好ましい態様において、その組成物は、疾患の処置のために宿主生体においてカテプシンKポリペプチドを発現するためのカテプシンKポリヌクレオチドを含む。この点に関して、特に好ましい態様は、内因性カテプシンKの異常活性に関連する機能不全の処置のための、または治療用に提供するための、ヒト患者における発現である。

【0014】本発明のその他の目的、特徴、優越点および見解は、以下の記載より当業者には明白であろう。しかしながら、以下の記載および具体的な実施例は本発明の好ましい態様を示すものであり、説明のためにのみ示すものであることを理解されたい。開示した本発明の意図および範囲内での種々の変化および改変は、以下の記載および本明細書のその他の部分の知識により、当業者に容易に明らかとなる。

【0015】定義

以下の説明は、本明細書、とりわけ実施例において汎用される幾つかの用語の理解を容易くするために示すものである。説明は便宜上示すものであり、本発明を限定するものではない。「DNA消化」とは、DNAの特定の配列でのみ作用する制限酵素による、DNAの触媒的切断を意味する。本明細書における種々の制限酵素は、市販により入手可能であり、反応条件、コファクターおよびそのたの使用に必要なものは、当業者に周知であり、通常的に用いられるものである。分析の目的では、典型的なものではプラスミドまたはDNA断片1 μ gを、反応緩衝液約20 ml中酵素約2単位で消化する。プラスミド構築用に、DNA断片を単離する目的では、典型的なものではDNA 5~50 μ gを、酵素約20~250単位で、相対的に高容量中で消化する。

【0016】特定の制限酵素のための適当な緩衝液および基質量は、以下に引用するような標準的な実験マニュアルに記載されており、それらは商業的供給者により規定されている。37℃で約1時間のインキュベーション時間は、通常用いられるものであるが、標準的な方法、供給者の指示書および反応の特色に応じて変更できる。消化後、当業者に通常の周知の方法により、反応物を分析し、寒天またはポリアクリルアミドゲルの電気泳動により断片を精製できる。

【0017】「遺伝的要素」とは、一般的にポリペプチドをコードする領域もしくは転写もしくは翻訳、もしくは宿主細胞においてポリペプチドを発現するのに重要なその他の過程を制御する領域を含むポリヌクレオチド、またはポリペプチドをコードする領域、およびそれらに機能的に連結した発現を制御する領域の両方を含むポリヌクレオチドを意味する。

【0018】遺伝的要素は、エピソーム要素即ち宿主細胞ゲノムと物理的に独立した分子として複製するベクター内に含まれる得る。これらは真核細胞におけるメソトレキセート選別によりトランスフェクトしたDNAの増幅中に現れるようなミニ染色体の中に含まれる得る。遺伝的要素はまた、天然の状態ではなく、むしろ単離、クローニングおよび宿主細胞への導入のような操作の後に、精製DNAの形態で宿主細胞ゲノム内またはとりわけベクター内にも含まれる得る。

【0019】「同一性」とは、ポリペプチドまたはポリヌクレオチド配列の2本の鎖の間の適合の同一性により決定できるような、二つのポリペプチドまたは二つのポリヌクレオチド配列間の配列相関性の程度を意味する。同一性は容易に算出できる。二つのポリヌクレオチドまたはポリペプチド配列間の同一性を測定する方法は数多くあるが、「同一性」なる用語は、当業者に周知である (Computational Molecular Biology, Lesk, A.M. 編、オックスフォードユニバーシティプレス、ニューヨーク、1988年; Biocomputing: Informatics and Genome Projects, Smith, D.W. 編、アカデミックプレス、ニューヨーク、1993年; Computer Analysis of Sequence Data, 第一部, Griffin, A.M. およびGriffin, H.G. 編、ヒューマンプレス、ニュージャージー、1994年; Sequence Analysis in Molecular Biology, von Heinje, G., アカデミックプレス、1987年; およびSequence Analysis Primer, Gribskov, M. およびDevereux, J. 編、Mストックトンプレス、ニューヨーク、1991年)。二つの配列を同一性を測定するのに用いる一般的な方法には、Guide to Huger Computers, Martin, J. Bishop編、アカデミックプレス、サンディエゴ、1994年並びにCarillo, H. およびLipman, D., SIAM J. Applied Math., 48: 1073 (1988) に開示されているもの等があるが、これらに限定するものではない。同一性を測定するための好ましい方法は、試験する二つの配列間の最も大きな適合部分を得るように設

計したものである。このような方法は、コンピュータープログラムに組み立てられている。二つの配列間の同一性を測定するための好ましいコンピュータープログラム法には、これらに限定するものではないが、GCGプログラムパッケージ (Devereux, J. ら、Nucleic Acids Research, 12(1):387 (1984))、BLASTP、BLASTN、FASTA (Atschul, S. F. ら、J. Molec. Biol. 215:403 (1990)) 等がある。

【0020】「単離」とは、「人工的に」天然の状態から変化させられた、すなわち、それが天然に存在する場合、元来の環境から変化させるもしくは取り出す、またはその両方を行ったことを意味する。例えば天然発生のポリヌクレオチドまたはポリペプチドが、天然の状態で生存動物に存在する場合は、「単離」されていないが、同一のポリヌクレオチドまたはポリペプチドが、天然に共存する物質の幾つかまたは全てから分離されている場合は、本明細書に用いる用語である、「単離」がなされている。単離の一部として、または単離に引き続いて、このようなポリヌクレオチドは、DNAの様なその他のポリヌクレオチドに結合して突然変異を誘発し、融合タンパク質を形成し、例えば宿主中で伸長または発現することができる。単離ポリヌクレオチドは単独で、またはベクターの様なその他のポリヌクレオチドと結合して、培養物中または全生物体中の宿主細胞に導入できる。培養物中または全生物体中の宿主細胞に導入されたDNAも本明細書に用いる用語である、単離をしたといえる。というのはこれらは天然発生の形態または環境にはないからである。同様に、ポリヌクレオチドおよびポリペプチドは、培地製剤例えば細胞にポリヌクレオチドまたはポリペプチドを導入するための溶液、化学的または酵素的反応のための組成物または溶液等の組成物にも生じることができ、これらは天然には存在しない組成物であり、本明細書に用いる用語の単離されたポリヌクレオチドまたはポリペプチドの範疇である。

【0021】「ライゲーション」とは、最も多くは2本鎖DNAである、2個またはそれ以上のポリヌクレオチド間のホスホジエステル結合を形成する工程を意味する。ライゲーション技術は当業界に周知であり、ライゲーションプロトコールは標準的な実験マニュアルおよび参考文献、例えばSambrookら、Molecular Cloning, A Laboratory Manual, 第2版; コールドスプリングハーバー・ラボラトリープレス、コールドスプリングハーバー、ニューヨーク (1989) およびManiatisら、(以下に引用) 146頁に記載されている。

【0022】「(複数の) オリゴヌクレオチド」とは、比較的短いポリヌクレオチドを意味する。この用語はしばしば一本鎖デオキシリボヌクレオチドを意味するが、同様に一本鎖、二本鎖または三本鎖リボヌクレオチド、アンチセンスポリヌクレオチド、RNA:DNAハイブリッドおよびとりわけ二本鎖DNAを意味することがで

きる。オリゴヌクレオチド、例えば一本鎖DNAプローブオリゴヌクレオチドは、しばしば自動オリゴヌクレオチド合成器にかける等の化学的方法により合成される。しかしながら、オリゴヌクレオチドはインビトロ組換えDNA媒介法および細胞および生体におけるDNAの発現によるなどの種々のその他の方法により作ることができる。

【0023】最初、化学合成したDNAは、典型的なものでは5'リン酸塩なしで得られる。このようなオリゴヌクレオチドの5'末端は、組換えDNA分子を形成するのに通常使用されるDNAリガーゼを用いるライゲーション反応によるホスホジエステル結合形成のための基質ではない。このようなオリゴヌクレオチドのライゲーションが望まれる場合、キナーゼおよびATPを用いるような標準法により、リン酸塩を添加できる。化学合成したオリゴヌクレオチドの3'末端は、通常遊離の水酸基を有し、T4DNAリガーゼ等のリガーゼの存在下、容易に別のオリゴヌクレオチドのような別のポリヌクレオチドの5'リン酸塩とホスホジエステル結合を形成する。周知のように、この反応は、望む場合、ライゲーションの前にその他の(複数の)ポリヌクレオチドの5'リン酸塩を除去することにより、選択的に防御できる。

【0024】「プラスミド」は、本明細書では通常前に小文字のpを付し、および/または続いて大文字および/または数字で示し、当業者に周知の標準命名法に準じる。本明細書に開示する出発プラスミドは、制約を受けずに市販により、公に入手可能であるかまたは、周知の確立された技法を通常的に適用することにより、入手可能なプラスミドから構築することができる。本発明に準じて使用できる多くのプラスミド並びにその他のクローニングおよび発現ベクターは、周知であり、当業者に容易に入手できる。さらに、当業者は本発明の使用に適したその他のプラスミドをいくらかでも容易に構築することができる。本発明におけるこのようなプラスミドおよびその他のベクターの特性、構築および使用は、本明細書より当業者に容易に理解され得る。

【0025】「(複数の) ポリヌクレオチド」とは、通常改変されていないRNAもしくはDNA、または改変されたRNAもしくはDNAになり得る任意のポリリボヌクレオチドまたはポリデオキシリボヌクレオチドを意味する。従って、例えば本明細書で用いるポリヌクレオチドは、とりわけ一本鎖および二本鎖DNA、一本鎖および二本鎖領域の混合物であるDNA、一本鎖および二本鎖RNA、並びに一本鎖および二本鎖領域の混合物であるRNA、一本鎖もしくはより通常的には二本鎖、または一本鎖および二本鎖領域の混合物でよいDNAおよびRNAを含むハイブリッド分子を意味する。さらに、本明細書で用いられるポリヌクレオチドは、RNAもしくはDNAまたはRNAとDNAの両方を含む三本鎖領域を意味する。そのような領域における鎖は同一の分子ま

たは異なる分子からのものでよい。この領域は一つまたはそれ以上の分子の全てを含み得るが、より典型的には分子の一部の領域のみを含む。三重らせんの分子の一つは、しばしばオリゴヌクレオチドである。

【0026】本明細書で用いるポリヌクレオチドなる用語には、一つまたはそれ以上の改変した塩基を含有する、上述のDNAまたはRNA等が含まれる。このように、安定性またはその他の理由で改変した骨格を有するDNAまたはRNAは、本明細書で意図する用語である「ポリヌクレオチド」である。さらに、イノシン等の通常のでない塩基、またはトリチル化塩基等の改変塩基を含むDNAまたはRNAは（二つの例だけをを示す）、本明細書で用いられる用語である、ポリヌクレオチドである。当業者に既知の多くの有用な目的を提供するDNAおよびRNAに、非常に多くの改変がなされていることは、明らかであろう。本明細書で用いられるポリヌクレオチドなる用語は、ポリヌクレオチドのこのような化学的、酵素的または代謝的に改変した形態、およびとりわけウイルスおよび単細胞および複合細胞等の細胞に特徴的なDNAおよびRNAの化学的形態を包含する。

【0027】本明細書で用いる「ポリペプチド」には、以下に記載する全てのポリペプチドを包含する。ポリペプチドの基本的な構造は周知であり、非常に多くの参考書およびその他の当業界の発行物に記載されている。これに鑑み、本明細書で用いるこの用語は、ペプチド結合により直鎖で互いに結合している2個またはそれ以上のアミノ酸を含む任意のペプチドまたはタンパク質を意味する。本明細書で用いるこの用語は、当業界で通常例えばペプチド、オリゴペプチドおよびオリゴマー、とも称する両方の短鎖、および多くの型がある当業界で一般的にタンパク質と称する長鎖を意味する。

【0028】ポリペプチドはしばしば通常20個の天然発生アミノ酸と称される20個のアミノ酸以外のアミノ酸を含有し、末端アミノ酸を含む多くのアミノ酸は、プロセッシングおよびその他の翻訳後改変のような天然の工程によるのみならず、また当業者に周知の化学的改変技術によっても、上記のポリペプチドを改変できることは理解されよう。ポリペプチドに天然に起こる通常的な改変は余りにも多いので、余すところなく揭示することはできないが、基礎的な参考書およびさらに詳細な論文並びに多数の研究文献にも詳しく記載されており、これらは当業者に周知である。本発明のポリペプチドに施すことができる既知の改変には、説明のために幾つか示すと、アセチル化、アシル化、ADP-リボシル化、アミド化、フラビンの共有結合、ヘム部分の共有結合、ヌクレオチドまたはヌクレオチド誘導体への共有結合、脂質または脂質誘導体の共有結合、ホスホチジイルノシトールへの共有結合、交差架橋、環状化、ジスルフィド結合形成、ジメチル化、交差架橋共有結合形成、システイン形成、ピログルタメート形成、ホルミル化、ガンマー-カ

ルボキシル化、グリコシル化、GPIアンカー形成、水酸化、ヨウ素化、メチル化、ミリスチル化、酸化、タンパク質加水分解プロセッシング、リン酸化、プレニル化、ラセミ化、セレノイル化、硫酸化、アルギニル化のようなトランスファーRNA媒介のタンパク質へのアミノ酸の添加、およびユビキチネーション (ubiquitination) などがある。

【0029】このような改変は当業者に周知であり、科学文献に非常に詳細に記載されている。幾つかのとりわけ通常的な改変、例えばグリコシル化、脂質結合、硫酸化、グルタミン酸残基のガンマー-カルボキシル化、水酸化およびADP-リボシル化等が最も基礎的な参考書、例えばProteins-Structure and Molecular Properties、第2版、T. E. Creighton、W. H. Freeman and Company、ニューヨーク（1993）に記載されている。多くの詳細な論文例えば、Posttranslational Covalent Modification of Proteins、B. C. Johnson編、アカデミックプレス、ニューヨーク（1983）のWold、F.、Posttranslational Protein Modifications : Perspective and Prospects、1~12頁；Seifterら、Analysis for Protein Modifications and nonprotein cofactors、Meth. Enzymol. 182:626-646（1990）およびRattanら、Protein Synthesis : Posttranslational Modification and Aging、Ann. N. Y. Acad. Sci. 663:48-62（1992）の記載が、本主題に利用できる。

【0030】ポリペプチドはいつも完全に直鎖であるとは限らないということは周知であり、上述の通りであると理解されよう。例えばポリペプチドはユビキチネーションの結果分岐でき、一般的には、天然のプロセッシングなどの翻訳後の事象および天然では起こり得ない人為的操作によりもたらされる事象の結果、分岐のある、または分岐のない環状にできる。環状、分岐および分岐した環状ポリペプチドは、非翻訳天然の工程により、および同様に全く合成的な方法で、合成できる。改変はペプチド骨格、アミノ酸側鎖およびアミノまたはカルボキシル末端など、ポリペプチドの至る所で起こりうる。実際、共有結合の改変による、ポリペプチドのアミノまたはカルボキシル基またはその両方の遮断は、天然に起こり、合成ポリペプチドおよびこのような改変は同様に本発明のポリペプチドに存在し得る。例えば、大腸菌 (E. coli.) で作られるポリペプチドのアミノ末端残基は、タンパク質溶解のプロセッシングの前に殆ど必ずN-ホルミルメチオニンになる。

【0031】ポリペプチドに起こる改変はしばしばそれの作り方に影響される。例えば宿主のクローン遺伝子の発現により作られるポリペプチドでは、大部分の改変の特性および程度は宿主細胞の翻訳後改変能力およびポリペプチドのアミノ酸配列に存在する改変信号により決定される。例えば、周知のように、グリコシル化はしばしば大腸菌のような細菌宿主では起こらない。従って、グリコシル化が望まれる場合、ポリペプチドはグリコシル

化宿主、一般的には真核細胞に発現すべきである。昆虫細胞がしばしば哺乳動物細胞と同じ翻訳後グリコシル化に用いられ、この理由で昆虫細胞発現系が、とりわけグリコシル化の本来のパターンを有する哺乳動物タンパク質を効率よく発現するように改造されている。同様の考察がその他の改変にも適用される。

【0032】改変の同一の型は、上記ポリペプチドの幾つかの部分で同じまたは異なる程度で存在しうると理解されよう。また、上記ポリペプチドには改変の多くの型が存在できる。一般的に、本明細書で用いる用語であるポリペプチドは、このような全ての改変特に宿主細胞においてポリヌクレオチドを発現することにより合成したポリペプチドに存在する改変を包含する。

【0033】本明細書で用いるポリヌクレオチドおよびポリペプチドの「(複数の) 変異型」なる用語は、参照ポリヌクレオチドまたはポリペプチドとは各々異なるポリヌクレオチドまたはポリペプチドである。この意味での変異型は、本明細書の以下およびその他の部分でより詳細に記載する。

(1) 別の配列、参照ポリヌクレオチドとはヌクレオチド配列が異なるポリヌクレオチド。一般的に差異はヌクレオチドの参照配列および変異型が、全体的に非常に類似しており、多くの領域で同等であるように限定される。以下に記すように、変異型におけるヌクレオチドの配列の変化はサイレントでよい。すなわち、ポリヌクレオチドによりコードされるアミノ酸には変化がなくてよい。変化がこの型のサイレント変化には限定される場合、変異型は、参照体と同一のアミノ酸配列を有するポリペプチドをコードするであろう。また以下に記すように、変異型のヌクレオチド配列における変化が、参照ポリヌクレオチドによりコードされるポリペプチドのアミノ酸配列を変化させ得る。このようなヌクレオチドの変化は、以下に論じるように、参照配列によりコードされるポリペプチドにおけるアミノ酸置換、添加、欠損、融合およびトランケーションを招く。

【0034】(2) アミノ酸配列が他すなわち参照ポリペプチドとは異なるポリペプチド。一般的に差異は、参照体および変異型の配列が、全体的に非常に類似しており、多くの領域で同等であるように限定される。変異型および参照ポリペプチドは、1またはそれ以上の置換、添加、欠損、融合およびトランケーションが任意の組み合わせで起こることにより、アミノ酸配列が変化する。

【0035】本明細書で用いる「受容体分子」とは、本発明のカテプシンKポリペプチドと特異的に結合または相互作用する分子であり、古典的な受容体および酵素基質(両者共に好ましい)のみならず本発明のポリペプチドと特異的に結合または相互作用するその他の分子(各々「結合分子」「相互作用分子」、並びに「カテプシンK結合分子」および「カテプシンK相互作用分子」とも

称する)も含まれる。これらの、カテプシンK結合分子はまた例えばカテプシンK基質類似体を含む。本発明のポリペプチドおよびこのような分子、例えば受容体または結合もしくは相互作用分子等との結合は、本発明のポリペプチドに対して独占的にでき、これは非常に好ましいことであり、または本発明のポリペプチドに非常に特異的にでき、これは非常に好ましいことであり、または本発明のポリペプチドを含むタンパク質群に非常に特異的にでき、これは非常に好ましいことであり、または本発明のポリペプチドを少なくとも一つ含むタンパク質群の幾つかに非常に特異的にできる。受容体、例えば本発明のポリペプチドに特異的に結合する抗体および抗体由来物質もまた非天然発生である。

【0036】

【発明の実施の形態】本発明はとりわけ以下により詳細に記載する、新規カテプシンKポリペプチドおよびポリヌクレオチドに関する。具体的には、本発明は、ウサギOC-2およびヒトカテプシンOcDNAに対してアミノ酸配列相同性という点で関連している新規ヒトカテプシンKのポリペプチドおよびポリヌクレオチドに関する。Tezuka, K.ら、J. Biol. Chem.、269:1106-1109 (1994)。本発明は特に図1-9【配列番号:1】に示すヌクレオチド配列を有するカテプシンK、および本明細書において「寄託クローン」または「寄託クローンのgDNA」と称するATCC寄託番号98035のgDNAのカテプシンKヌクレオチド配列に関する。図1-9

【配列番号:1】に示すヌクレオチド配列は、本明細書の別の部分でさらに具体的に示すように、寄託クローンのgDNAをシーケンシングすることにより入手したことは理解されよう。従って、寄託クローンの配列は、二者間の任意の不一致に関しては、寄託クローンの配列に従うものとする。

【0037】ポリヌクレオチド

本発明の一つの態様に準じて、図10および11【配列番号:20】の誘導アミノ酸配列(誘導アミノ酸配列に関しては図28をも参照)を有するカテプシンKポリペプチドをコードする単離ポリヌクレオチド、または寄託クローンのgDNAによりコードされるカテプシンKポリペプチドを提供する。本明細書に提供する情報を用いて、図1-9【配列番号:1】に示すポリヌクレオチド配列のようなヒトカテプシンKポリペプチドをコードする本発明のポリヌクレオチドは、出発物質としてヒトの細胞からのDNAを用いて、gDNAをクローニングするような標準的なクローニングおよびスクリーニングの手法を用いて得ることができる。本発明の実例となる図1-9【配列番号:1】に示すポリヌクレオチドは、実施例1に記載するヒトgDNAライブラリーで見出された。

【0038】寄託クローンのヒトカテプシンKをコードするgDNAシーケンシングの結果に示されるよう

に、本発明のヒトカテプシンKはカテプシンファミリーのその他のタンパク質に構造的に相関する。このように得られたgDNA配列を図1-9【配列番号：1】に示す。これはイントロンは除去するが、エキソンは全て含む、約329個のアミノ酸基のタンパク質をコードする非近接的なオープンリーディングフレームを含有する。本発明のポリヌクレオチドは、クローニングにより得るか、もしくは化学合成技術により産生するか、またはそれらを組み合わせて得られた、mRNAまたはhnRNAなどのRNAの形態で、または例えばcDNAおよびgDNAなどのDNAの形態でよい。DNAは三本鎖、二本鎖または一本鎖でよい。一本鎖DNAは有義（センス）鎖として知られているコード鎖でよく、またアンチセンス鎖とも称する非コード鎖でもよい。ポリペプチドをコードするコード配列は、図1-9【配列番号：1】に示すポリヌクレオチドまたは寄託クローンのポリヌクレオチドのエキソン配列に同等にできる。ポリヌクレオチドは、また、遺伝子コードの冗長性（退化）の結果、図10および11【配列番号：20】のDNAまたは寄託gDNAのポリペプチドをコードする異なる配列を有するポリペプチド、例えばこれらに限定するものではないが、このようなgDNAから転写したスプライス変異型等でもよい。

【0039】図1-9【配列番号：1】のポリペプチドまたは寄託gDNAによりコードされるポリペプチドをコードする本発明のポリヌクレオチドには、限定するものではないが、単独の成熟ポリペプチドのコード配列；成熟ポリペプチドをコードする配列および付加コード配列、例えばブレ、プロ、プレプロタンパク質配列等のリーダーまたは分泌配列をコードするコード配列；限定するものではないが、例えば転写、例えばスプライシングおよびポリアデニル化信号等のmRNAプロセッシング、リボソーム結合およびmRNAの安定性において役割を担う、転写された非翻訳配列等のイントロン並びに非コード5'および3'配列等の、付加非コード配列と共に、前述の付加コード配列を伴うかまたは伴わない成熟ポリペプチドのコード配列；付加機能性を付与するような付加アミノ酸をコードする付加コード配列等がある。このように、例えばポリペプチドはペプチドのような誘導ポリペプチドの精製を促すマーカ配列例えばペプチドを融合できる。本発明のある好ましい態様において、マーカ配列はヘキサ-ヒスチジンペプチド、例えば、とりわけベクターpQE-9に提供されるタグであり、多くが市販により入手可能である。例えばGentzら、Proc. Natl. Acad. Sci., USA, 86:821-824 (1989)に記載されるように、ヘキサ-ヒスチジンは融合タンパク質の精製に便宜的に提供される。HAタグはインフルエンザヘマググリンタンパク質由来のエピトープに対応し、これについては例えばWilsonら、Cell, 37:767 (1984)に記載されている。

【0040】前述に準じ、本明細書で用いる「ポリペプチドをコードするポリヌクレオチド」なる用語は、本発明のポリペプチド、とりわけ図10および11【配列番号：20】に示すアミノ酸配列を有するヒトカテプシンK、または寄託クローンのgDNAによりコードされるヒトカテプシンKのアミノ酸配列をコードする配列などのポリヌクレオチドを包含する。この用語は、（例えばイントロンに干渉される）ポリペプチドをコードする単一の連続領域または不連続領域を付加領域と共に含み、コードおよび/または非コード配列をも含有し得るポリヌクレオチドを包含する。

【0041】本発明はさらに、図10および11【配列番号：20】の誘導アミノ酸配列を有するポリペプチド、または寄託クローンのgDNAのエキソンによりコードされるポリペプチドの断片、類似体および誘導体をコードする、上述のポリヌクレオチドの変異型に関し、これに限定するものではないが、このようなgDNAから転写したスプライス変異型等がある。ポリヌクレオチドの変異型には天然発生アレル変異型またはスプライス変異型のような天然発生変異型でよく、または天然に発生することが知られていない変異型でもよい。このような非天然発生のポリヌクレオチドの変異型は、ポリヌクレオチド、細胞または生体等に突然変異誘発技法を施したりすることにより作ることができる。このようなポリヌクレオチドの非天然発生変異型は、スプライス受体、供与体および/または分岐部位を改変することにより、またはgDNAを天然には発現しない細胞またはこのような細胞から作った細胞抽出物中にgDNAを発現することにより作ることができる。この観点における変異型は、前述のヌクレオチド置換、欠損または付加によるポリヌクレオチドとは異なる。置換、欠損または付加は1またはそれ以上のヌクレオチドに起こる。変異型はコードもしくは非コード領域またはその両方において変化し得る。コード領域における変化は保存的または非保存的アミノ酸置換、欠損または付加を産み出す。

【0042】この点に関する本発明の特に好ましい態様は、図1-9【配列番号：1】に示すカテプシンKのポリヌクレオチド配列、または寄託クローンのgDNAのカテプシンKのポリヌクレオチド配列；それらの変異型、類似体、誘導体および断片、並びに変異型、類似体および誘導体の断片である。この点に関する本発明の特に好ましい態様は、カテプシンK変異型、類似体、誘導体および断片、並びにその断片の変異型、類似体および誘導体の断片をコードするポリヌクレオチドであり、図10および11【配列番号：20】のまたは寄託物のカテプシンKポリペプチドのアミノ酸配列に幾つか、少しの、5~10、1~5、1~3、2、1または0個のアミノ酸残基を置換、欠損または付加を任意の組み合わせで施したアミノ酸配列を有する。中でもとりわけ好ましいものは、サイレント置換、付加および欠損であり、カ

テブシンKの特性および活性を変化しない。また、この点に関してとりわけ好ましいものは、保存置換である。最も好ましいものは、置換を施さない、図10および11【配列番号：20】の、または寄託物のアミノ酸配列を有するポリペプチドである。

【0043】本発明のさらに好ましい態様は、図10および11【配列番号：20】に示すアミノ酸配列を有するカテブシンKをコードするポリヌクレオチド、または上述のそれらの変異型、正確な相同体、誘導体および類似体、並びにこのようなポリヌクレオチドに相補的なポリヌクレオチドに、少なくとも70%の同一性を有するポリヌクレオチドである。また別に、最も好ましいポリヌクレオチドは寄託クロンのgDNAのカテブシンKポリペプチドをコードするポリヌクレオチド、およびそれらに相補的なポリヌクレオチドに少なくとも80%の同一性を有する領域を含むポリヌクレオチドである。この点に関して、とりわけ好ましいポリヌクレオチドは少なくとも90%の同一性を有するものであり、特に好ましいのは少なくとも95%の同一性を有するものである。さらに少なくとも95%の同一性を有するものの中でも少なくとも97%であるのがより好ましく、中でも少なくとも98%および少なくとも99%であるのが特に好ましく、さらに少なくとも99%であるのがより好ましい。

【0044】本発明のさらに好ましい態様は、カテブシンKイントロンポリヌクレオチド配列を包含するポリヌクレオチド、特にイントロン1【配列番号：4】、2【配列番号：6】、3【配列番号：8】、4【配列番号：10】、5【配列番号：12】、6【配列番号：14】または7【配列番号：16】を包含し、図1-9【配列番号：1】および12-24【配列番号：2-19】に示す配列を有するポリヌクレオチド、または上述のそれらの変異型、正確な相同体、誘導体および類似体、並びにこのようなポリヌクレオチドに相補的なポリヌクレオチドである。本発明のその他の好ましい態様は、カテブシンKイントロン1【配列番号：4】、2【配列番号：6】、3【配列番号：8】、4【配列番号：10】、5【配列番号：12】、6【配列番号：14】または7【配列番号：16】を包含し、カテブシンK以外の遺伝子のエクソンに効果的に連結された、またはカテブシンKエクソンおよびその他の遺伝子のエクソンを結合したポリヌクレオチドである。

【0045】また本発明のその他の好ましい態様は、カテブシンKエクソンポリヌクレオチド配列を含むポリヌクレオチド、特に、エクソン1【配列番号：3】、2【配列番号：5】、3【配列番号：7】、4【配列番号：9】、5【配列番号：11】、6【配列番号：13】、7【配列番号：15】または8【配列番号：17】を包含し、図1-9【配列番号：1】および12-24【配列番号：2-19】に示すエクソンポリヌクレ

オチド配列を有するポリヌクレオチド、または上述のそれらの変異型、正確な相同体、誘導体および類似体、並びにこのようなポリヌクレオチドに相補的なポリヌクレオチドである。本発明のその他の好ましい態様は、カテブシンKエクソン1【配列番号：3】、2【配列番号：5】、3【配列番号：7】、4【配列番号：9】、5【配列番号：11】、6【配列番号：13】、7【配列番号：15】または8【配列番号：17】を包含し、カテブシンK以外の遺伝子のイントロンに効果的に連結されたポリヌクレオチドである。

【0046】本発明のさらに好ましい態様は、特異的にスプライスしたポリヌクレオチド、特に以下のエクソン-エクソンペア：1-3、1-4、1-5、1-6、1-7、1-8、2-4、2-5、2-6、2-7、2-8、3-4、3-5、3-6、3-7、3-8、4-5、4-6、4-7、4-8、5-7、5-8または6-8を何れか一つまたはそれ以上含むポリヌクレオチドである。本発明の特に好ましい態様は、細胞内で機能するポリペプチド、特にカテブシンKの生物学的活性を有するポリペプチド、とりわけヒト細胞に発現するポリペプチドをコードする特異的にスプライスしたポリヌクレオチドである。

【0047】エクソン-エクソンペアを含むポリヌクレオチドは、天然発生のスプライス変異型のような天然発生の変異型でよく、または天然に発生すると知られていない変異型でよい。このようなポリヌクレオチドの非天然発生の変異型は突然変異誘発技術をこれらをポリヌクレオチド、細胞または生体に適用したりして作ることができる。このようなポリヌクレオチドの非天然発生の変異型は、スプライス受体、供与体および/または分岐部位を改変することにより、または天然にはgDNAを発現しない細胞またはこのような細胞から作った細胞抽出物中でgDNAを発現させることにより作ることができる。エクソン-エクソンペアは完全に融合したエクソンか、またはエクソンの断片を、存在するスプライス結合部で融合できる。エクソン断片を含む好ましいエクソン-エクソンペアは、一つが機能的なスプライス供与体部位を含み、他方が機能的なスプライス受体部位を含み、両者がイントロンにより機能的に連結している、少なくとも2つのエクソンから作ることができる。

【0048】この点に関するとりわけ好ましい態様は、図10および11【配列番号：20】のcDNAまたは寄託クロンのgDNAにコードされる成熟ポリペプチドと実質的に同一の生物学的機能または活性を保持するポリペプチドをコードするポリヌクレオチドである。本発明はさらに本明細書上述の配列にハイブリダイズするポリヌクレオチドに関する。この点に関して、本発明は特にストリンジェント状態で本明細書上述のポリヌクレオチドにハイブリダイズするポリヌクレオチドに関する。本明細書で用いる「ストリンジェント状態」なる用語は、配列間の同一性が少なくとも95%、好ましくは

少なくとも97%である場合に起こるハイブリダイゼーションを意味する。

【0049】本発明のポリヌクレオチド検定に関してここでさらに論じるが、例えば上述の本発明のポリヌクレオチドは、カテプシンKをコードするcDNA全長およびゲノムクローンを単離するため、およびヒトカテプシンK遺伝子に高度な配列類似性を有するその他の遺伝子のcDNAおよびゲノムクローンを単離するためのcDNAおよびゲノムDNAのハイブリダイゼーションプローブとして使用することができる。このようなプローブは通常少なくとも15塩基を含む。好ましくはこのようなプローブは少なくとも30塩基を有し、少なくとも50塩基を有することができる。とりわけ好ましいプローブは少なくとも30塩基を有し、50塩基以下である。例えばカテプシンK遺伝子のコード領域は、既知DNA配列を用いてオリゴヌクレオチドプローブを合成し、スクリーニングすることにより単離できる。本発明の遺伝子の配列に相補的な配列を有する標識オリゴヌクレオチドは、次に、ヒトcDNA、ゲノムDNAまたはmRNAのライブラリーのスクリーニングに用い、プローブがハイブリダイズするライブラリーのメンバーを決定する。

【0050】本発明のポリヌクレオチドおよびポリペプチドは、ヒト疾患の処置法および診断法の発見のために研究物質および材料として使用でき、とりわけポリヌクレオチド検定に関連して本明細書でさらに論じる。本発明のポリヌクレオチドは付加アミノ酸またはカルボキシル末端アミノ酸を加えた成熟タンパク質、または成熟タンパク質に内在するアミノ酸である（例えば成熟形態が一つ以上のポリペプチド鎖を有する場合）ポリペプチドをコードできる。このような配列は前駆体から成熟形態へのタンパク質のプロセッシングに役割を担い、タンパク質のトラフィッキング (trafficking) を促進し、タンパク質の半減期を延長もしくは短縮し、またはとりわけ検定もしくは産生のためのタンパク質の操作を容易にすることができる。一般的には本来、付加アミノ酸は、細胞酵素によりプロセッシングされ、成熟タンパク質から取り除かれる。

【0051】1またはそれ以上のプロ配列と融合したポリペプチドの成熟形態を有する前駆タンパク質は、ポリペプチドの不活性形態にできる。プロ配列が除去されると、このような不活性前駆体が通常活性化される。プロ配列のいくつかまたは全ては、活性化の前に除去できる。通常、このような前駆体はプロタンパク質と称される。要するに、本発明のポリヌクレオチドは成熟タンパク質、リーダー配列を加えた成熟タンパク質（プレタンパク質と称することができる）、プレタンパク質のリーダー配列ではない1またはそれ以上のプロ配列を有する成熟タンパク質の前駆体、またはリーダー配列および1またはそれ以上のプロ配列を有するプロタンパク質の前駆体であるプレプロタンパク質であり、プロ配列は通常

ポリペプチドの活性および成熟形態を産み出すプロセッシング段階で除去される。

【0052】寄託物

ヒトカテプシンK gDNAを含有する寄託物は、上述のアメリカン・タイプ・カルチャー・コレクションに寄託した。また上述のとおり、gDNA寄託物は本明細書では「寄託クローン」または「寄託クローンのgDNA」と称する。寄託クローンは、アメリカン・タイプ・カルチャー・コレクション、12301パークローンドライブ、ロックビル、メリーランド20852、米国に1996年4月26日に寄託し、ATCC寄託番号98035を付与された。寄託物は、カテプシンK gDNAの全長を含有するP1コスミドであり、寄託時に「P1SacB2CatK/P129」と称される。

【0053】寄託は、特許手続き上の微生物寄託の国際承認に関するブダペスト条約の条件下で為されている。特許が発行されると何らの制限または条件もなく、最終的には分譲される。寄託は当業者の便宜のためにのみ提供され、35 U. S. C. 112条のもとに要求されるような、寄託が実施可能要件であることを承認するものではない。本明細書の配列に関する任意の記載に矛盾する事象は、寄託物に含まれるポリヌクレオチド配列およびそれによりコードされるポリペプチドのアミノ酸配列に従う。寄託物を製造、使用または販売するためにはライセンスが必要であるが、そのようなライセンスはここでは賦与されていない。

【0054】ポリペプチド

本発明はさらに図10および11【配列番号：20】の誘導アミノ酸配列を有するヒトカテプシンKポリペプチドに関し、これは図1-9【配列番号：1】の配列から転写した、スプライスしていないかまたは特異的にスプライスしたhnRNAまたはmRNAによりコードされるか、または寄託クローンによりコードされるアミノ酸配列を有する。また、ミスセンスまたはノンセンス突然変異を含むカテプシンK gDNAによりコードされるポリペプチド、または少なくとも一つのイントロンをさらに含む、非スプライスまたは部分的にスプライスしたhnRNAによりコードされるこれらのポリペプチド、とりわけ天然に細胞、特にヒトの細胞に見出されるこれらのポリペプチドをも提供する。フレームシフト突然変異は疾患と関連することが示されている (Hol, FA.ら、Journal of Medical Genetics, 32(1)52-56 (1995))。

【0055】本発明が提供する好ましいポリペプチドは、特異的にスプライスしたポリヌクレオチドによりコードされ、とりわけ以下のエキソン-エキソンペア：1-3、1-4、1-5、1-6、1-7、1-8、2-4、2-5、2-6、2-7、2-8、3-4、3-5、3-6、3-7、3-8、4-5、4-6、4-7、4-8、5-7、5-8または6-8を何れか一つまたはそれ以上含むポリヌクレオチドにコードされるポリペプチドである。本発明の特に好ましい態様

は、特異的にスプライスしたポリヌクレオチドによりコードされるポリペプチドであり、ポリペプチドが細胞内で機能し、特にカテプシンKの生物学的活性を有し、さらにはヒト細胞内に発現するポリペプチドである。

【0056】本発明のさらに好ましい態様は、エキソンポリヌクレオチド配列を含むポリヌクレオチド、特にカテプシンKエキソン1 [配列番号：3]、2 [配列番号：5]、3 [配列番号：7]、4 [配列番号：9]、5 [配列番号：11]、6 [配列番号：13]、7 [配列番号：15] または8 [配列番号：17] を包含し、図1-9 [配列番号：1] および12-24 [配列番号：2-19] に示すエキソンポリヌクレオチド配列を有するポリヌクレオチド、または上述のそれらの変異型、正確な相同体、誘導体および類似体、並びにこのようなポリヌクレオチドに相補的なポリヌクレオチドによりコードされるポリペプチドである。本発明のその他の好ましい態様は、カテプシンKエキソン、エキソン1 [配列番号：3]、2 [配列番号：5]、3 [配列番号：7]、4 [配列番号：9]、5 [配列番号：11]、6 [配列番号：13]、7 [配列番号：15] または8 [配列番号：17] を包含し、カテプシンK以外の遺伝子のイントロンに機能的に連結したまたは他の遺伝子のエクソンに結合したポリヌクレオチドによりコードされるポリペプチドである。

【0057】本発明はまたこれらのポリペプチドの断片、類似体および誘導体にも関する。「断片」、「誘導体」および「類似体」なる用語は図10および11 [配列番号：20] のポリペプチド、図1-9 [配列番号：1] の配列から転写される、スプライスされないまたは特異的にスプライスされたhnRNAまたはmRNAによりコードされるポリペプチド、または寄託gDNAによりコードされるポリペプチドを称している場合、このようなポリペプチドと本質的に同一の生物学的機能または活性を保持するポリペプチドを意味する。従って、類似体にはプロタンパク質部分が切断されて活性成熟ポリペプチドを産生して活性化できるプロタンパク質等がある。本発明のポリペプチドは組換えポリペプチド、天然ポリペプチドまたは合成ポリペプチドでよい。特定の好ましい態様では、これは組換えポリペプチドである。

【0058】図10および11 [配列番号：20] のポリペプチド、または図1-9 [配列番号：1] の配列から転写される、非スプライスまたは特異的にスプライスされたhnRNAもしくはmRNAによりコードされるポリペプチド、または寄託クローンのgDNAによりコードされるポリペプチドの断片、誘導体もしくは類似体は、(i) 1またはそれ以上のアミノ酸残基が保存または非保存アミノ酸残基（好ましくは保存アミノ酸残基）により置換されており、このような置換アミノ酸残基は遺伝的コードによりコードされたポリペプチドかコードされていないポリペプチド、(ii) 1またはそれ以上の

アミノ酸残基が置換基を含むポリペプチド、または(ii) 成熟ポリペプチドがその他の化合物例えばポリペプチドの半減期を伸ばす化合物（例えばポリエチレングリコール）と融合したポリペプチド、または(iv) 付加アミノ酸が成熟ポリペプチド、例えばリーダーもしくは分泌配列、または成熟ポリペプチドもしくはプロタンパク質配列の精製に用いられる配列と融合するポリペプチドでよい。このような断片、誘導体および類似体は本明細書の教示より当業者の範疇であると考えられる。

【0059】この点に関するとりわけ好ましい態様は、図10および11 [配列番号：20] に示すカテプシンKのアミノ酸配列を有するポリペプチド、それらの変異型、類似体、誘導体および断片、並びにその断片の変異型、類似体および誘導体である。また別に、この点に関するとりわけ好ましい態様は、寄託クローンのgDNAのカテプシンKのアミノ酸配列を有するポリペプチド、それらの変異型、類似体、誘導体および断片、並びにその断片の変異型、類似体および誘導体である。

【0060】とりわけ好ましい変異型は、保存アミノ酸置換により参照とは異なる変異型である。このような置換体は、ポリペプチドの上記アミノ酸を特性の類似した別のアミノ酸で置換したものである。典型的なものでは、保存置換でみられるのは脂肪族アミノ酸Ala、LeuおよびIle間での相互の置き換え；水酸基SerおよびThrの交換、酸性基AspおよびGluの取り替え、アミド基AsnおよびGln間での置換、塩基性基LysおよびArgの取り替えおよび芳香性基Phe、Tyr間での置き換えである。

【0061】この点に関するさらに好ましい態様は、図10および11 [配列番号：20] のカテプシンKポリペプチドのアミノ酸配列または寄託クローンのgDNAのアミノ酸配列を有し、幾つか、少しの、5~10、1~5、1~3、2、1または0個のアミノ酸残基を置換、欠損または付加を任意に組み合わせて施した変異型、類似体、誘導体および断片、並びにその断片の変異型、類似体および誘導体である。これらの中でもとりわけ好ましいのは、カテプシンKの特性および活性を変化させないサイレント置換、付加および欠損である。またこの点に関するとりわけ好ましい態様は、保存置換である。最も好ましいのは、置換していない図10および11 [配列番号：20] または寄託クローンのアミノ酸配列を有するポリペプチドである。本発明のポリペプチドおよびポリヌクレオチドは単離形態で提供されるのが好ましく、均質になるまで精製するのが好ましい。

【0062】本発明のポリペプチドには、配列番号：3、配列番号：5、配列番号：7、配列番号：9、配列番号：11、配列番号：13、配列番号：15または配列番号：17のエキソンの少なくとも一つのエキソンによりコードされるポリペプチド（とりわけ成熟ポリペプチド）、および配列番号：3、配列番号：5、配列番号：

号：7、配列番号：9、配列番号：11、配列番号：13、配列番号：15または配列番号：17のエキシンの少なくとも一つのエキシンのコードされるポリペプチドに少なくとも70%の類似性（好ましくは少なくとも70%の同一性）を有するポリペプチド、さらに好ましくは配列番号：3、配列番号：5、配列番号：7、配列番号：9、配列番号：11、配列番号：13、配列番号：15または配列番号：17のエキシンの少なくとも一つのエキシンのコードされるポリペプチドに少なくとも90%の類似性（さらに好ましくは少なくとも90%の同一性）を有するポリペプチド、なおさらに好ましくは配列番号：3、配列番号：5、配列番号：7、配列番号：9、配列番号：11、配列番号：13、配列番号：15または配列番号：17のエキシンの少なくとも一つのエキシンのコードされるポリペプチドに少なくとも95%の類似性（なおさらに好ましくは少なくとも95%の同一性）を有するポリペプチドなどがあり、また通常少なくとも30個のアミノ酸、さらに好ましくは少なくとも50個のアミノ酸を含むこのようなポリペプチド部分を有するこのようなポリペプチド部分などもある。

【0063】当業界に周知であるように、二つのポリペプチド間の「類似性」は一つのポリペプチドのアミノ酸配列およびその保存アミノ酸置換体を二つ目のポリペプチドの配列と比較して決定する。本発明のポリペプチドの断片または部分は、ペプチド合成により対応するポリペプチド全長を産生するのに用いることができる：従って、断片はポリペプチドの全長を産生するための媒介物質として用いることができる。本発明のポリヌクレオチドの断片または部分は本発明のポリヌクレオチド全長を合成するために用いることができる。

【0064】断片

また本発明の好ましい態様には、カテプシンKの断片、特に図1-9【配列番号：1】に示すエキソンによりコードされるアミノ酸、または寄託クローンのカテプシンKのエキソン配列によりコードされるアミノ酸を有するカテプシンKの断片、および図1-9【配列番号：1】または寄託クローンのカテプシンKのエキソン断片または変異型および誘導体等がある。この点に関して、断片は前述のカテプシンKポリペプチドおよびそれらの変異型または誘導体の全てではなく一部が全く同一であるアミノ酸配列を有するポリペプチドである。

【0065】このような断片は「フリースタンディング」すなわちその他のアミノ酸またはポリペプチド、例えばエキソンの一部かそれに融合したものではなく、一部もしくは領域を形成するより大きなポリペプチド内に含まれてもよい。より大きなポリペプチド内に含まれる場合、本明細書で論じる断片は単一の連続した領域を形成するのが最も好ましい。しかしながら、幾つかの断片は単一の、より大きなポリペプチド内に含まれ得る。例えば、特定の好ましい態様は、カテプシンK断片のアミ

ノ末端に融合した異型のブレおよびプロポリペプチド領域、および断片のカルボキシル末端に融合した付加領域を有する、宿主中に発現するように設計した前駆体ポリペプチドに含まれる本発明のカテプシンKポリペプチドの断片に関する。従って、本明細書の意図する一つの態様における断片は、カテプシンK由来の融合ポリペプチドまたは融合タンパク質の一つのまたは複数の部分を意味する。

【0066】本発明のポリペプチド断片の代表的な例としては、各々図1-9【配列番号：1】および12-24【配列番号：2-19】に示すエキソンまたはイントロン1、2、3、4、5、6または7ポリヌクレオチド配列を有する、カテプシンKエキソン1、2、3、4、5、6、7または8を含むポリヌクレオチド配列によりコードされる上述のポリペプチド断片、または上述のそれらの変異型、正確な相同体、誘導体および類似体、並びにこのようなポリヌクレオチドに相補的なポリヌクレオチドによりコードされるポリペプチドでよい。これに関連して、どちらかの末端または両方の末端の幾つかの、少しの、5、4、3、2または1個のアミノ酸により特に記載した、より大きいまたはより小さい範囲および複数の範囲を含める。例えば、これに関連して、約65-90個のアミノ酸は、65プラスまたはマイナス幾つかの、少しの、5、4、3、2または1個のアミノ酸、から90プラスまたはマイナス幾つかの、少しの、5、4、3、2または1個のアミノ酸、すなわち広くは65マイナス幾つかのアミノ酸から90プラス幾つかのアミノ酸、狭くは65プラス幾つかのアミノ酸から90マイナス幾つかのアミノ酸の範囲にできることを意味する。

【0067】この点に関して、より好ましくは、どちらかまたは両方の末端で5個ほどのアミノ酸をプラスまたはマイナスした上述の範囲である。とりわけ好ましくは、上述の末端のどちらかまたは両方で3個ほどのアミノ酸をプラスまたはマイナスした上述の範囲である。とりわけ好ましくは、どちらかまたは両方の末端で1個のアミノ酸をプラスまたはマイナスした範囲、または付加または欠損のない上述の範囲である。この点に関して、とりわけ最も好ましくは、カテプシンKのエキシンの各々によりコードされる断片である。

【0068】本発明のとりわけ好ましい断片は、カテプシンKのトランケーション突然変異である。トランケーション突然変異体には、アミノ末端を含む残基の一連の連続（すなわち連続領域、部分または部位）もしくはカルボキシル末端を含む残基の一連の連続の欠損、または二重トランケーション突然変異のように二つの残基の一連の連続（一つはアミノ末端を含み、一つはカルボキシル末端を含む）の欠損を除いた、図1-9【配列番号：1】または寄託クローンのエキソンによりコードされるアミノ酸配列を有するカテプシンKポリペプチド、また

はそれらの変異型もしくは誘導体などがある。提示した大きさの範囲の断片もまたトランケーション断片の好ましい態様であり、これは通常の断片の中でもとりわけ好ましい。

【0069】また本発明の好ましい態様は、カテプシンKの構造的または機能的属性により特徴づけられた断片である。この点に関する本発明の好ましい態様には、カテプシンKのアルファヘリックスおよびアルファヘリックス形成領域（「アルファ領域」）、ベータシートおよびベータシート形成領域（「ベータ領域」）、ターンおよびターン形成領域（「ターン領域」）、コイルおよびコイル形成領域（「コイル領域」）、親水性領域、疎水性領域、アルファ両極性領域、ベータ両極性領域、可変領域、表面形成領域および高抗原性指標領域を含む断片などがある。

【0070】特定の好ましい領域には、ガルニアールロブソナルファ領域、ベータ領域、ターン領域およびコイル領域、チョウファスマンアルファ領域、ベータ領域およびターン領域、カイトーデューリトル親水性領域および疎水性領域、アイゼンベルグアルファおよびベータ両極性領域、カルプラスーシュルツ可変領域、エミニ表面形成領域並びにジェームソンウルフ高抗原性指標領域などがある。

【0071】この点に関する非常に好ましい断片は、上述の幾つかの様相のような幾つかの構造的様相をくみあわせたカテプシンKの領域を含む断片である。この点に関して、図1-9【配列番号：1】のエキソン配列は、全てターン領域、親水性領域、可変領域、表面形成領域および高抗原性指標領域に非常に特徴的なアミノ酸組成物をコードする特性があり、特に好ましい領域である。このような領域は、大きなポリペプチドないに包含され得、単独で上述の本発明の好ましい断片と成りうる。このパラグラフで用いる「約」なる用語は、上述の断片に関して示す一般的な意味である。

【0072】さらに好ましい領域は、カテプシンKの活性を媒介する領域である。この点に関して最も好ましくは、カテプシンKの化学的、生物学的、抗原的またはその他の活性を有する断片であり、同様の活性もしくは改善された活性のある、または望ましくない活性を減じた断片である。本発明はまたとりわけ前述の断片をコードするポリヌクレオチド、その断片をコードするポリヌクレオチドにハイブリダイズしたポリヌクレオチド、特にストリンジェント状態でハイブリダイズするポリヌクレオチド、およびその断片をコードするポリヌクレオチドを増幅するPCRプライマーのようなポリヌクレオチドに関する。この点に関して、好ましいポリヌクレオチドは、上述の様に、好ましい断片に対応するポリヌクレオチドである。

【0073】その他の好ましいポリヌクレオチドは、カテプシンKの遺伝的要素であり、ポリアデニル化領域、

エンハンサー、プロモーター、キャップ部位イントロン、エキソンおよびスプライス部位等（これらの事象を記載する参考文献には以下のものがある；Darnel, J., ら, Molecular Cell Biology, 第2版, W.H. Freeman, ニューヨーク (1990) ; Watson, J.D. ら, Molecular Biology of the Gene, Benjamin/Cummings Pub., メンロパーク, カリフォルニア州 (1987)）があるが、これらに限定するものではない。

【0074】非翻訳領域には、遺伝子発現を制御するのに重要な多くの要素を含む。これらの領域の突然変異およびマーカーは疾患にも関連している (Ozawa T. ら, European Journal of Immunogenetics, APR, 22(2)163-169 (1995))。本発明の好ましい態様は、5' UTR、とりわけ図12 (A)【配列番号：2】に示す配列である。5' UTRにおける突然変異およびマーカーは疾患に関連している (Carlock, L. ら, Human Genetics, APR, 93(4)457-459 (1994))。とりわけ好ましいポリヌクレオチドはカテプシンK gDNAの5' UTR領域におけるエンハンサーおよびプロモーターである。エンハンサーはしばしば5' UTR領域に見出され、遺伝子発現を上方制御する（プロモーターに関する一般的な参考文献としてはMillerら, Biotechniques 7 : 980-990 (1989) 参照）。本発明のエンハンサーは異型遺伝子に遺伝子発現を上向制御（アップレギュレーション）するように機能的に融合できる。エンハンサープロモーターは組織特異性遺伝子発現を制御すると考えられ、とりわけ遺伝子発現に有用と考えられるものは破骨細胞および白血球、とりわけマクロファージ細胞である。特に好ましいポリヌクレオチドは図12 (A)【配列番号：2】の前述の配列を有するエンハンサープロモーターである。転写因子はしばしばエンハンサーおよびプロモーターと関与し、これらの領域の機能およびこれらの因子の結合部位を変調する作用については以下に記載されている (Faisst, SteffenおよびMeyer, Silke, Nucleic Acids Research, 20(1)3-26 (1991) ; Smale, Stephen T., Transcription: Mechanisms and Regulation, ラーベンプレスリミテッド63-81頁 (1994))。これらの部位は、例えば転写開始に関与するSp1、Ap1およびAp3のような因子に結合する (Faisst, SteffenおよびMeyer, Silke, Nucleic Acids Research, 20(1)3-26 (1991))。転写因子の標準的な結合部位は、図23-24 (S)【配列番号：2】の下線を付したところである。図23-24 (S)【配列番号：2】のPuボックスは、マクロファージすなわち遺伝子カテプシンKもまた見出される細胞に存在することが記載されている (Zhang, Dong-Er, Nol. and Cell. Biol. 14(1)373-381 (1994))。本発明はとりわけ破骨細胞および白血球、特にマクロファージにおける組織特異的発現の媒介に有用なプロモーター領域を提供する。エンハンサーおよびプロモーター領域に存在するPuボックス (AGGAA) もまたマクロファージセ

ルライン (THP1) において観察される。本発明の配列におけるPuボックスを提供する。これらのPuボックスはマクロファージ中のカテプシンK遺伝子中で活性であると考えられる。THP1細胞において行われるRT-PCRは、プローブとしてカテプシンK配列を用い、発現させる。プロモーターはカテプシンK遺伝子発現の制御の研究に、とりわけ疾患の診断のためのプローブとする領域として特に有用である。ビタミンD応答要素は、既知遺伝子の5' UTRに見出された (Kahlen, Jean-Pierre, およびCarlberg, Carsten, Biochemical & Biophysical Research Communications, 202(3):1366-1372 (1994) ; Darwish, Hisham およびDeLuca, Hector, Critical Reviews in Eukaryotic Gene Expression, 3(2): 89-116 (1993) ; Carlberg, Carsten, Eur. J. Biochem., 231:517-524 (1995) ; Ohyama, Yoshihiko, J. Biol. Chem., 269(14):10545-10550 (1994))。ビタミンD反応要素 (「vDハーフサイト」) およびカルシウムイオン (「Caハーフペア」) 反応要素の部分は図23-24 (S) [配列番号: 2] に示す3' UTR配列に存在する。このような部位は以下に記載されている: Katz, Ronald, W., Subauste, Jose, S. およびKoenig, Ronald J., J. Biol. Chem., 270(10):5238-5242 (1995)。図12 (A) および図23-24 (S) [配列番号: 2] に示す5' UTR配列に存在するその他のハーフサイトには、オステオポンチン/上皮小体ホルモン反応要素、カルシトロール反応要素およびオステオカルシンハーフサイト等がある (例えばJuge-Aubry, Christianaら, J. Biol. Chem., 270(30):18117-18122 (1995) 参照)。プロモーターおよびエンハンサー領域に見出され、本発明に提供されるプロモーター因子結合部位もまたまたカテプシンKイントロンに見出される。エストロゲン反応要素もまたカテプシンK 5' UTRに存在すると推定される。当業者は本明細書に提供される方法を用いて、これらの要素を容易に見出すことができる。

【0075】さらに好ましいポリヌクレオチドは(図10および11)に示す配列のATG出発コドンの上方49塩基対に位置するキャップサイト (cap site) である。本発明のさらに好ましい態様は、カテプシンKのプロモーター領域 (図12 (A) および図23-24 (S) [配列番号: 2]) である。機能的なプロモーター領域配列については記載されている (Corden, J. ら, Science, 209:1406-1414 (1990))。図12 (A) [配列番号: 2] および図23-24 (S) [配列番号: 2] に示すカテプシンK配列における非標準プロモーター領域には、出発コドンATGの上方19-27塩基対のA-Tに富む配列を含む。プロモーターのTATAボックス領域の突然変異は疾患と関連していることが示されている (Peltoketo, H. ら, Genomics, 23(1):250-252 (1994))。

【0076】カテプシンKの3' 非翻訳領域は、本発明

の好ましいポリヌクレオチドであり、とりわけ図22 (Q) [配列番号: 18] に示すポリヌクレオチド、特に図22 (R) [配列番号: 19] に示す領域が好ましい。3' UTRにおける突然変異は疾患と関連がある (Saito, A. ら, Journal of the American Society of Nephrology, 4(9):1649-1653 (1994) ; Payne, S.J. ら, Human Molecular Genetics, 3(2):390 (1994))。図22 (Q) [配列番号: 18] に示すポリアデニル化領域は、また好ましい3' UTRのポリヌクレオチド配列である。ポリアデニル化領域は標準的なポリアデニル化ヘキサヌクレオチドAATAAAの二つのコピーを包含する。ポリアデニル化領域は、例えば発現ベクター中にmRNA 3' 末端形成を媒介するために用いることができる (例えばGil, A. ら, Nature, 312:473-474 (ロンドン) (1984))。

【0077】本発明のその他のとりわけ好ましいポリヌクレオチドは、スプライス供与体、スプライス受体およびスプライス分岐点などのスプライス部位であるが、これらに限定するものではない。スプライス接合部形成はオープンリーディングフレームを適切に作るのに不可欠である (Mount, Stephen, M., Department of Molecular Biophysics and Biochemistry, エール大学、スターリングホールオブメディスン、ニューヘブレン、コネチカット州、米国、IRLプレスリミテッド、ロンドン459-472頁 (1981))。スプライス接合の不適切な形成に関連する疾患が知られている。とりわけ好ましいスプライス接合ポリヌクレオチドは図25に示す。

【0078】イントロンは遺伝子発現および成熟mRNAの形成に重要な要素を含む。イントロンの突然変異およびマーカーは疾患に関連することが示されている (Peral, G. ら, Human Molecular Genetics, APR4(4):569-574 (1995) ; Chrysogelos, S.A., Nucleic Acids Research, 21(24):5736-5741 (1993) ; Ameis, D., Journal of Lipid Research, 36(2):241-250 (1995))。スプライス接合もまた疾患に関連することが示されている (Ameis, D. ら, Journal of Lipid Research, 36(2):241-250 (1995年2月) ; Petrini, J.H. ら, Journal of Immunology, 152(1):176-183 (1994) ; Kleiman, F.B. ら, Human Genetics, 94(3):279-282 (1994))。また別にスプライシングおよび未解明のスプライス部位選別もまた疾患に関連することが示されている (Arakawa, H. ら, Human Molecular Genetics 3(4):565-568 (1994) ; Tieu, P.T. ら, Human Mutation, 3(3):333-336 (1994) ; Reale, M.A. ら, Cancer Research, 54(16):4493-4501 (1994))。イントロンはまたこれらの配列の一部としてエンハンサー要素を含むことができる。本発明の好ましい態様は、カテプシンKイントロンとりわけ図13-21 (C, E, G, I, K, MおよびO) [配列番号: 4, 6, 8, 10, 14および16] に示す配列を有するカテプシンKイントロンである。イントロンの多形性は連鎖分析後に疾患

のマーカ―として提供できる。さらに本明細書に記載する遺伝的分析は、疾患に関連したおおよび／または疾患の原因となるイントロンにおける突然変異を示すために用いることができる。

【0079】その他の好ましい態様はカテプシンKイントロンエンハンサーである。イントロン3はスプライス接合GT/AG則に一致しない。このイントロン/エキソン境界はP1クローンおよびゲノムDNAのシーケンシングにより証明された。GC/AGスプライス接合は一般的ではないが、以下に記載されている；Mount, Stephen, M., Department of Molecular Biophysics and Biochemistry, エール大学、スターリングホールオブメディスン、ニューヘブレン、コネチカット州、米国、IRLプレスリミテッド、ロンドン459-472頁（1981）。本発明のさらに好ましい態様は、カテプシンKエキソンととりわけ図13-21（B、D、F、H、J、L、NおよびP）【各々配列番号：3、5、7、9、11、13、15および17】に示す配列を有するこれらのエキソンである。エキソンの多形性は連鎖分析後の疾患のマーカ―として提供できる。さらに本明細書に記載する遺伝的分析は、疾患に関連するおおよび／または疾患の原因となるエキソンにおける突然変異を示すために用いることができる。

【0080】本発明のポリヌクレオチド断片は、カテプシンK遺伝子の発現を阻害するリボザイム（ribozyme）を作るのに用いることができる。リボザイム構築のための一般的構築法は当業界で周知である（Stram, Y. および Molad, T., Virus Genes, 9(2):155-159（1995））。当業者は本発明の新規断片を用いるこれらの方法を容易に適用し、新規リボザイムを構築するために、容易に適用することができる。好ましいリボザイム構築物にはカテプシンK遺伝子の転写された制御要素に相補的な配列、とりわけ5' 非翻訳領域、スプライス接合部および3' 非翻訳領域、とりわけボリアデニル化領域に相補的なポリヌクレオチドを含む。

【0081】本発明の断片、とりわけ非翻訳領域における領域、プロモーターおよびイントロンは、疾患とりわけ骨疾患例えば骨粗鬆症、および例えばバジネット病、ゴーチャーズ病、CNS炎症、アルツハイマー病、上皮小体亢進症、骨分解、転移性腫瘍、リウマチ性関節症、変形性関節症、歯周病および骨インプラントおよび骨プロシーゼス（protheses）とりわけ歯科インプラントの分解に診断プローブとして有用である。さらに、疾患のマーカ―は、本発明の診断法に有用である、カテプシン遺伝子の領域、とりわけ非翻訳領域において示すことができる。

【0082】ベクター、宿主細胞、発現

本発明はまた、本発明のポリヌクレオチドを含むベクター、本発明のベクターで遺伝子操作する宿主細胞および組換え技法による本発明のポリペプチドの産生にも関す

る。宿主細胞は、遺伝子操作して、ポリヌクレオチドを挿入し、本発明のポリペプチドを発現することができる。例えば感染、形質導入、トランスフェクション、トランスベクションおよび形質転換の周知の技法を用いて、ポリヌクレオチドを宿主細胞に導入できる。ポリヌクレオチドは単独でまたは他のポリヌクレオチドと共に導入できる。このような他のポリヌクレオチドは別個に導入、同時導入または本発明のポリヌクレオチドに結合させて導入できる。従って、例えば哺乳動物細胞で同時トランスフェクションおよび選別するための標準的な技法を用いて、選別マーカ―をコードする別の離れたポリヌクレオチドコードと共に、本発明のポリヌクレオチドを宿主細胞にトランスフェクトできる。この場合、ポリヌクレオチドは通常遺伝子的に安定して宿主細胞ゲノムに挿入される。

【0083】また別に、ポリヌクレオチドは宿主細胞における伸長のための選別マーカ―を含有するベクターに結合できる。ベクター構築は上述の技法により、宿主細胞に導入できる。一般的に、プラスミドベクターは沈殿例えばリン酸カルシウム沈殿中、または荷電脂質との複合体中DNAとして導入する。ポリヌクレオチドを宿主に導入するために、電気泳動をも用いることができる。ベクターがウイルスである場合、インビトロでパッケージングまたはパッケージング細胞に導入でき、パッケージングウイルスを細胞に形質導入できる。本発明の態様に準じた、ポリヌクレオチドの製造および細胞へのポリヌクレオチドの導入に適した種々の技法は、当業者に周知であり、日常的に用いられるものである。このような技法は上述のSambrookらに詳細に報告されており、これらの技法を詳細に説明する多くの実験マニュアルの説明となるものである。

【0084】本発明のこの態様に準じたベクターは、例えばプラスミドベクター、一本または二本鎖ファージベクター、一本または二本鎖RNAまたはDNAウイルスベクターなどでよい。このようなベクターはポリヌクレオチド、好ましくはDNAとして、DNAおよびRNAの細胞への導入のための周知の技術により細胞に導入できる。ベクターがファージおよびウイルスベクターである場合もまた、感染および形質導入のため周知の技術により、パッケージ化または包膜化したウイルスとして細胞に導入でき、これは好ましいことである。ウイルスベクターは複製能力があるかまたは複製能力を欠損している。後者の場合のウイルスの増殖は通常宿主細胞を完全化する場合にのみ起こる。中でも好ましいベクターは、ある観点では、本発明のポリヌクレオチドおよびポリペプチドを発現するベクターである。一般に、このようなベクターは、発現させるポリヌクレオチドに機能的に連鎖した宿主における発現に有効なシス作用制御領域を含む。適当なトランス作用因子は宿主により供給される、完全化ベクターにより供給される、または宿主への

導入時にベクター自身により供給されるかのいずれかである。

【0085】この点に関するある態様では、ベクターは特異的発現を提供する。このような特異的発現は、誘導可能な発現もしくは特定の型の細胞においてのみ発現するか、または誘導可能および細胞特異性の両方でよい。誘導可能なベクターのうちとりわけ好ましいベクターは、温度および栄養添加物など操作が容易である環境因子により発現を誘導できるベクターである。本発明のこの態様に適した種々ベクターには原核および真核宿主で用いられる構成および誘導可能な発現ベクター等があり、当業者に周知であり、通常的に用いられている。操作した宿主細胞は、従来の栄養培地で培養でき、とりわけ活性化プロモーター、選別形質転換体または増幅遺伝子に適当なように改変できる。発現のために選別した宿主で以前に用いた培養条件例えば温度、pH等は、当業者に明白であるように、通常本発明のポリペプチドの発現に適している。

【0086】非常に多くの種類のベクターが、本発明のポリペプチドを発現するために使用できる。このようなベクターには、染色体、エピソームおよびウイルス由来のベクター、例えば細菌プラスミド由来、バクテリオファージ由来、酵母エピソーム由来、酵母染色体要素由来、例えばバキュロウイルス、パポウイルス、例えばSV40、ワクシナウイルス、アデノウイルス、鶏痘ウイルス、仮性狂犬病ウイルスおよびレトロウイルス等のウイルス由来のベクター、並びにそれらを組み合わせた物に由来するベクター、例えばプラスミドおよびバクテリオファージ遺伝的要素由来のベクター、例えばコスミドおよびファージミド (phagemids) 等があり、全て本発明の態様に準じた発現に用いることができる。通常宿主中にポリペプチドを発現するためのポリヌクレオチドを保持、伸長または発現するのに適した任意のベクターが、この点に関して発現に使用できる。

【0087】周知のおよび日常的な種々の任意の技術により、適当なDNA配列をベクターに挿入できる。一般に、1またはそれ以上の制限エンドヌクレアーゼでDNA配列および発現ベクターを切断し、T4 DNAリガーゼを用いて制限断片を結合することにより、発現のためのDNA配列を発現ベクターに結合する。この目標のために用いることができる制限およびライゲーション方法は、当業者に周知であり、日常的に用いられているものである。この点に関して、および別法を用いた構築発現ベクターのための適切な方法もまた、当業者に周知であり、日常的に用いられているものであり、本明細書に引用したSambrookら、に非常に詳細に記載されている。

【0088】発現ベクターのDNA配列は、適当な(複数の)発現制御配列、例えばmRNA転写を指示するプロモーター等に機能的に連結する。このようなプロモーターの代表的なものとしては、幾つか周知のプロモーター

の名前を挙げると、ファージラムダPLプロモーター、大腸菌 (*E. coli*) lac、trpおよびtacプロモーター、SV40初期および後期プロモーター並びにレトロウイルスLTRのプロモーター等がある。本発明の態様に用いるのに適した、記載していない多くのプロモーターが周知であり、本明細書の記載および実施例に説明する方法で当業者に容易に用いることができるということは、理解されよう。

【0089】一般に、発現構築物は転写開始および終了部位を含み、転写された領域では翻訳のためのリボソーム結合部位を含む。構築物により発現される成熟転写物のコード部分には、翻訳されるポリペプチドの末端に適当に位置する開始および終了コドンで翻訳開始AUGを含む。さらに構築物は、発現を制御および引き起こす制御領域を含有してよい。通常、多くの一般的な慣習的な方法に準じて、このような領域は制御転写例えばとりわけレプレッサー結合部位およびエンハンサーにより操作される。

【0090】伸長および発現のためのベクターは通常選別可能なマーカーを含有する。このようなマーカーはまた増幅にも適してよいが、またはベクターがこの目的のために更なるマーカーを含有できる。この点に関して、発現ベクターは、形質転換した宿主細胞の選別のための表現型特性を提供する、1またはそれ以上の選別マーカーを含有する。好ましいマーカーにはジヒドロフオレトリダクターゼまたは真核細胞培養用ネオマイシン抵抗、大腸菌およびその他の細菌の培養用テトラサイクリン、カナマイシン、およびアンピシリン抵抗遺伝子等がある。

【0091】本明細書記載の適当なDNA配列を含有するベクター、および適当なプロモーター、およびその他の適当な制御配列は、望ましいポリペプチドをそこに発現するのに適した種々の周知の技術を用いて適当な宿主に導入できる。適当な宿主の代表的な例としては、細菌細胞例えば大腸菌、ストレプトミセスおよびネズミチフス菌 (*Salmonella typhimurium*) 細胞; 真菌細胞例えば酵母細胞; 昆虫細胞例えばドロソフィラS2 (*Drosophila S2*) およびスポドプテラSf9 (*Spodoptera Sf9*) 細胞; 動物細胞例えばCHO, COSおよびボウズ (Bos) 黒色腫細胞; 並びに植物細胞等がある。非常に多様な発現構築のための宿主が周知であり、本明細書の開示により、本発明の態様に準じたポリペプチドを発現するための宿主の選択が、当業者に容易になる。

【0092】さらに、本発明はまた上述の配列を1またはそれ以上含む、組換え構築物例えば発現構築物も含む。構築物には、本発明のこのような配列が挿入されているベクター、例えばプラスミドまたはウイルスプラスミド等も含まれる。この配列は正または逆配向で挿入できる。この点に関して、ある好ましい態様において、構築物には、さらにその配列に機能的に連結した制御配列

例えばプロモーター等がある。非常に多くの適当なベクターおよびプロモーターが当業者に周知であり、本発明の使用に適した多くのベクターが市販により入手可能である。

【0093】市販されている以下のベクターを実施例として提供する。細菌中で使用するのに好ましいベクターには、キアゲンより入手可能なpQE70、pQE60およびpQE-9；ストラックジーンより入手可能なpBSベクター、ファジェスクリプト (Phagescript) ベクター、ブルースクリプト (Bluescript) ベクター、pNH8A、pNH16a、pNH18AおよびpNH46A；並びにファルマシアより入手可能なptc99a、pKK223-3、pKK233-3、pDR540、pRIT5等がある。真核細胞ベクターで好ましいものは、ストラックジーンより入手可能なpWLNEO、pSV2CAT、pOG44、pXT1およびpSG；並びにファルマシアより入手可能なpSVK3、pBPV、pMSGおよびpSVL等がある。これらのベクターは単に多くの市販のベクターの説明のために列挙しており、本発明の態様に準じた使用のために、当業者に入手可能なベクターである。宿主内で本発明のポリヌクレオチドまたはポリペプチドを例えば導入、保持、伸長または発現するのに適した任意のその他のプラスミドまたはベクターが、本発明のこの態様において使用できることは理解されよう。

【0094】プロモーター領域は、プロモーター領域を欠いたリポーター転写ユニット、例えばクロラムフェニコールアセチルトランスフェラーゼ (「CAT」) 転写ユニットを、制限部位または志願プロモーター (candidate promoter) 断片すなわちプロモーターを含み得る断片を導入する部位の下方に含有するベクターを用いて、任意の望ましい遺伝子から選別できる。周知の様に、cat遺伝子の上方の制限部位でプロモーター含有断片をベクターに導入すると、標準的なCAT検定により検出できるCAT活性の産生を引き起こす。この目標に適したベクターは周知であり、容易に入手可能である。2個のこのようなベクターはpKK232-8およびpCM7である。このように本発明のポリヌクレオチドの発現のためのプロモーターには、周知で容易に入手できるプロモーターのみならず、リポーター遺伝子を用いて前述の技術により容易に得ることができるプロモーターもある。

【0095】本発明の好ましい態様は、プロモーターとして機能するカテプシンKプロモーター配列を含む発現ベクターである。このようなベクター構築物は、カテプシンKプロモーターを利用する細胞例えば破骨細胞およびマクロファージ中の標的遺伝子発現に用いることができる。目的の任意の遺伝子はカテプシンKプロモーターに発現可能なように連結し、カテプシンKプロモーターを利用するこのような細胞中に発現できる。この方法で

初代真核細胞例えばSV40T-抗原を不滅化する遺伝子を、カテプシンKプロモーターに発現可能なように連結でき、細胞例えば破骨細胞およびマクロファージ等の骨細胞を不滅化できる。特定の好ましいベクターは毒素遺伝子例えばリシンに発現可能なように連結したカテプシンKプロモーターを含み、遺伝子発現にカテプシンKプロモーターを利用する細胞群を標的にして死滅させる方法に有用である。その他の好ましいベクターは、抗カテプシンKリボザイムまたはアンチセンスポリヌクレオチドに発現可能なように連結したカテプシンKプロモーターを含み、このような標的化死滅方法に有用である。

【0096】本発明に準じたポリヌクレオチドおよびポリペプチドの発現に適した、周知の細菌性プロモーターには、大腸菌lacIおよびlacZプロモーター、T3およびT7プロモーター、gptプロモーター、ラムダPR、PLプロモーター並びにtrpプロモーター等がある。この点に関して適当な周知の真核細胞プロモーターには、CMV即時型プロモーター、HSVチミジンキナーゼプロモーター、初期および後期SV40プロモーター、レトロウイルスLTRのプロモーター、例えばラウス肉腫ウイルス (「RSV」) のプロモーター並びにメタロチオネインプロモーター例えばマウスメタロチオネイン-Iプロモーター等がある。宿主細胞における発現のための適当なベクターおよびプロモーターの選別は周知の方法であり、発現ベクター構築、宿主へのベクターの導入および宿主中での発現に必要な技術は当業者に通常的に用いられるものである。

【0097】本発明はまた上述の構築物を含有する宿主細胞にも関する。宿主細胞は高等な真核細胞、例えば哺乳動物細胞もしくは昆虫細胞、もしくは下等な真核細胞、例えば酵母細胞でよく、または宿主細胞は原核細胞例えば細菌細胞でもよい。構築物の宿主細胞への導入はリン酸カルシウムトランスフェクション、DEAE-デキストラン媒介トランスフェクション、カチオン脂質媒介トランスフェクション、電気泳動、形質導入、感染またはその他の方法により影響を受け得る。このような方法は多くの標準的な実験マニュアル、例えばDavisら、Basic Methods in Molecular Biology, (1986) に記載されている。

【0098】宿主内の構築物は通常の方法で用いて、組換え配列によりコードされる遺伝子産物を産生することができる。また別に、本発明のポリペプチドは通常のペプチド合成により合成的に産生できる。成熟タンパク質は哺乳動物細胞、酵母、細菌またはその他の細胞に、適当なプロモーターの制御下で発現できる。細胞フリーの翻訳系は本発明のDNA構築物に由来するRNAを用いたこのようなタンパク質の産生にも用いることができる。原核および真核細胞宿主での使用に適当なクローニングおよび発現ベクターは、Sambrookら、Molecular Cloning: A Laboratory Manual, 第二編、コールドスプ

リングハーバーラボラトリープレス、コールドスプリングハーバー、ニューヨーク州(1989)に記載されている。

【0099】一般的に、酵母の組換え発現ベクターは複製源、下方の構造配列の転写を指示する高度に発現した遺伝子由来のプロモーターおよびベクターに曝露した後の細胞を含有するベクターの単離を可能にする選別マーカールなどを含む。なかでも適切なプロモーターには、グルコース分解酵素例えば、とりわけ3-ホスホグリセレートキナーゼ(「PGK」) α -因子、酸ホスファターゼおよび熱ショックタンパク質等をコードする遺伝子に由来するプロモーター等がある。選別マーカールには、大腸菌のアンピシリン抵抗遺伝子およびビール酵母菌の *trp1* 遺伝子等がある。高等真核細胞による本発明のポリペプチドをコードするDNAの転写は、エンハンサー配列をベクターに挿入することにより増強させることができる。エンハンサーは、与えられた宿主細胞型におけるプロモーターの転写活性を増強するように作用する、通常約10~300塩基対(bp)のDNAのシス作用要素である。エンハンサーの例としては、100~270塩基対での複製起源の後位置に位置するSV40エンハンサー、サイトメガロウイルス初期プロモーターエンハンサー、複製起源の後位置のポリオーマエンハンサー、およびアデノウイルスエンハンサー等がある。

【0100】本発明のポリペプチドの異型の構造配列をコードする本発明のポリヌクレオチドは、通常標準法を用いてベクターに挿入し、発現用プロモーターに機能的に連結する。ポリヌクレオチドは、転写開始部位がリボソーム結合部位の5'に適当に存在するように位置決定する。リボソーム結合部位は発現すべきポリペプチドの翻訳を開始するAUGの5'である。通常、開始コドン、通常AUGで始まるオープンリーディングフレームは他になく、リボソーム結合部位および開始AUGの間にある。また、通常、ポリペプチドの末端に翻訳停止コドンがあり、転写された領域の3'末端に適当に配置されたポリアデニル化信号および転写終了信号がある。

【0101】翻訳タンパク質を、小胞体内腔、ペリプラスミックスペースまたは細胞外環境へ分泌するために、適当な分泌信号を発現したポリペプチドに挿入することができる。信号はポリペプチドに対して内因性でよく、または異型性の信号でもよい。ポリペプチドは改変した形態、例えば融合タンパク質の形態で発現でき、分泌信号のみならず付加的な異型性機能的領域をも含み得る。このように、例えば付加アミノ酸とりわけ荷電アミノ酸領域は、ポリペプチドのN-末端に付加し、精製中またはそれに続く操作および保存中の、宿主細胞内での安定性および持続性を改善する。また、その領域はポリペプチドに付加して精製を容易にできる。このような領域はポリペプチドの最終的な調製の前に除去できる。ペプチド部分をポリペプチドに付加し、とりわけ、分泌または

排泄を引き起こし、安定性を改善し、および精製を容易にするのは、当業界でよくあることであり、通常の技術である。

【0102】本発明に準じたポリヌクレオチドおよびポリペプチドの伸長、保持または発現に適した原核細胞宿主には、大腸菌(*Escherichia coli*)、枯草菌(*Bacillus subtilis*)およびネズミチフス菌(*Salmonella typhimurium*)等がある。シュドモナス属、ストレプトミセス属、およびブドウ球菌属は、この点に関して、適切な宿主である。さらに、多くのその他の宿主も当業者に周知であり、使用できる。細菌使用のための有用な発現ベクターの代表的なものとして、周知のクローニングベクターpBR322(ATCC37017)の遺伝的要素を含む市販されているプラスミド由来選別マーカールおよび細菌の複製起源等があるが、これに限定するものではない。このような市販されているベクターには、例えばpKK2233(ファルマシアファインケミカルズ、ウプサラ、スウェーデン)およびGEM1(プロメガバイオテック、メディソン、ウイスコンシン州、米国)等がある。これらのpBR322「骨格」部分は適当なプロモーターおよび構造配列と結合して発現される。

【0103】適当な宿主株の形質転換および適当な細胞密度までの宿主株の成長に続いて、選択したプロモーターが誘導可能な場合、適当な手段(例えば温度シフトまたは化学的インデューサーへの曝露)によりそのプロモーターを誘導し、細胞は更なる期間培養する。通常、次に細胞を遠心により収穫し、物理学的または化学的手段により分解し、得られた粗抽出物をさらなる精製用に保持する。タンパク質の発現に用いた微生物細胞は、凍結-融解サイクル、ソニケーション、機械的分解または細胞溶解物質の使用等の、任意の通常的な方法により分解でき、このような方法は当業者に周知である。

【0104】種々の哺乳動物細胞系も同様に発現に用いることができる。哺乳動物発現系の例としては、サル腎臓繊維芽細胞COS-7ライン等があり、Gluzmanら、Cell, 23:175(1981)に記載されている。適合するベクターを発現できるその他のセルラインには、例えばC127、3T3、CHO、HeLa、ヒト腎293およびBHKセルライン等がある。哺乳動物発現ベクターには、複製起源、適当なプロモーターおよびエンハンサー等があり、また発現に必要な、任意の必須リボソーム結合部位、ポリアデニル化部位、スプライス供与部位および受容部位、転写終止配列および5'フランキング非転写配列などもある。この点に関する好ましい態様において、SV40スプライス部位およびSV40ポリアデニル化部位由来のDNA配列はこれらの型の獲得非転写遺伝要素に用いられる。

【0105】カテプシンKポリペプチドは硫酸アンモニウムまたはエタノール沈殿、酸抽出、陰イオンまたは陽イオン交換クロマトグラフィー、ホスホセルロースクロ

マトグラフィー、疎水性相互作用性クロマトグラフィー、親和性クロマトグラフィー、ヒドロキシアパタイトクロマトグラフィーおよびレクチンクロマトグラフィー等の周知の方法により、組換え細胞培養物から回収および精製できる。最も好ましくは、高速液体クロマトグラフィー（「HPLC」）を精製に用いる。ポリペプチドが単離および／または精製中に変性した場合、再び活性な立体配座にするために再折り畳みタンパク質のための周知の技術を用いることができる。

【0106】本発明のポリペプチドには、天然に精製された産物、化学合成法の産物および組換え法により原核または真核宿主、例えば細菌、酵母、高等植物、昆虫および哺乳動物細胞から産生した産物等がある。組換え産生法において用いる宿主に応じて、本発明のポリペプチドはグリコシル化または非グリコシル化してよい。さらに、本発明のポリペプチドはまた、ある場合は宿主媒介工程の結果として、開始改変メチオニン残基をも含有できる場合もある。

【0107】本発明の好ましい態様の更なる説明
カテプシンKポリヌクレオチドおよびポリペプチドは本発明に準じて、種々の適用ととりわけカテプシンKの化学的および生物学的特性を利用する適用のために使用できる。なかでも、とりわけ骨粗鬆症等の骨疾患、並びに例えばバジエット病、ゴーチャーズ病、CNS炎症、アルツハイマー病、上皮小体亢進症、骨分解、転移性腫瘍、リウマチ性関節症、変形性関節症、歯周病および骨インプラントおよび骨プロシーゼス（protheses）とりわけ歯科インプラントの分解等の疾患の検出および処置に適用される。さらに細胞、組織および器官の障害の診断および処置にも適用される。これらの本発明の態様は、さらに以下の記載により説明する。

【0108】ポリヌクレオチド検定
本発明はまた、例えば診断用試薬として相補的ポリヌクレオチドを検出するためのカテプシンKエキソン、イントロン、プロモーターおよびポリヌクレオチドの使用に関する。機能不全に関与するカテプシンKの突然変異型の検出は、例えば骨粗鬆症、歯周病、バジエット病、ゴーチャーズ病、CNS炎症、アルツハイマー病、上皮小体亢進症、および骨分解、転移性腫瘍、並びに骨インプラントおよび骨プロシーゼス、とりわけ歯科インプラントの分解等のカテプシンKの発現不足、過剰発現または変化した発現の結果起こる疾患または疾患に対する感受性を付与できる、または明確にできる診断方法を提供する。

【0109】ヒトカテプシン遺伝子の突然変異を担うものは種々の技術によりDNAレベルで検出できる。診断用の核酸は患者の細胞例えば血液、尿、唾液、組織生検および剖検材料より得ることができる。ゲノムDNAは直接的に検出するのに使用できるか、または分析の前にPCRを用いることにより酵素的に増幅できる（Saiki

ら、Nature, 324:163-166（1986））。ライゲーション媒介増幅もまた増幅に用いることができる（Vollach, V. ら、Nucl. Acids Res., 22:2507（1994））。RNAまたはcDNAもまた同じ方法で用いることができる。例としては、カテプシンKをコードする核酸に相補的なPCRプライマーはカテプシンK発現および突然変異の同定および分析に用いることができる。例えば、欠損および挿入は正常の遺伝子型に比較して増幅産物の大きさの変化により検出できる。点突然変異は、増幅DNAを放射性標識化カテプシンK RNAに、また別に放射性標識化カテプシンKアンチセンスDNA配列にハイブリダイズすることにより同定できる。完全にマッチした配列はRNアーゼ消化により、または融解温度の差により、ミスマッチ二重らせんから区別できる。

【0110】参照遺伝子および突然変異を有する遺伝子の配列の違いはまた、直接的なDNAシーケンシングにより表すこともできる。さらに、クローン化DNA部分は特異的DNA部分を検出するためのプローブとして用いることができる。このような方法の感度はPCRを適当に使用することにより、または別の増幅方法により非常に増強させることができる。例えば、シーケンシングプライマーは二本鎖PCR産物または改変PCRにより生じる一本鎖鋳型分子と共に用いられる。配列決定は、放射性標識化ヌクレオチドを用いる通常の方法により、または蛍光タグを用いる自動シーケンシング法により行うことができる。

【0111】DNA配列の差に基づく遺伝的試験は、変性物質を伴うまたは伴わないゲル中のDNA断片の電気泳動の可動性の変化を検出することにより達成できる。小さな配列の欠損および挿入は高分解能ゲル電気泳動により可視化できる。異なる配列のDNA断片は、異なるDNA断片の可動性がゲル中特異的融解または部分的融解温度に準じて異なる位置で遅延する、変性ホルムアミドグラジエントゲル上で区別できる（例えばMeyersら、Science, 230:1242（1985）参照）。

【0112】特異的な位置での配列の変化はまた、ヌクレアーゼ保護検定例えばRNアーゼおよびS1保護または化学的切断法によっても表すことができる（例えばCottonら、Proc. Natl. Acad. Sci., 米国, 85:4397-4401（1988））。このように、特異的DNA配列の検出は、ハイブリダイゼーション、RNアーゼ保護、化学的切断、直接的DNAシーケンシングまたは制限酵素の使用（例えば制限断片長多形性（「RFLP」））、SSCPおよびゲノムDNAのサザンブロッティングにより達成できる。通常のゲル電気泳動およびDNAシーケンシングに加えて、突然変異もまた従来の分析により検出できる。

【0113】染色体検定

本発明の配列はまた染色体同定にも有用である。この配列は個々のヒト染色体上の特定の位置を特異的に標的化

し、ハイブリダイズできる。さらに、染色体上の特定の部位を同定する必要性が現在ある。今のところ染色体位置をマークできる、実際の配列データ（繰り返し多形性）に基づく染色体マーキング試薬は少ししかない。本発明に準じた染色体に対するDNAのマッピングは、これらの配列を疾患に関連する遺伝子と関連付けるのに重要な最初の段階である。

【0114】この点に関する好ましい態様において、本明細書に開示するgDNAは、カテプシンK遺伝子のゲノムDNAをクローンするのに用いられる。これは種々の周知の技術および通常市販されているライブラリーを用いて達成できる。次にゲノムDNAを、この目的のための周知の技術を用いて、従来の染色体マッピングに使用する。染色体マッピングのための通常的方法に準じて、典型的なものでは、従来の良好なハイブリダイゼーション信号を提供するゲノムプローブを同定するための幾つかの試行錯誤が必要である。

【0115】さらに、gDNAからのPCRプライマー（好ましくは15～25塩基対）を調製することにより配列が染色体にマッピングされ得る場合もある。遺伝子の3' 非翻訳領域のコンピューター分析は、増幅法を複雑化する、ゲノムDNAにおいて1以上のエクソンを延ばさないプライマーを速やかに選別するために使用する。これらのプライマーは次に個々のヒト染色体を含有する体細胞ハイブリッドのPCRスクリーニングに使用される。プライマーに対応するヒト遺伝子を含有するこれらのハイブリッドのみが増幅断片を生じる。

【0116】体細胞ハイブリッドのPCRマッピングは、特定のDNAを特定の染色体に割り当てるための迅速な方法である。同一のオリゴヌクレオチドプライマーを用いる本発明を使用して、類似の方法で特異的染色体（例えば放射ハイブリッドパネル）または大きなゲノムクロンのプールからの、断片のパネルで亜局在化（sublocalize）できる。その染色体にマッピングするために使用できるその他のマッピング方法には、従来のハイブリダイゼーション、標識化流動分類染色体でのプレスクリーニングおよび構築染色体特異的cDNAライブラリーに対するハイブリダイズによるプレ選別等がある。

【0117】分裂中期染色体分散に対する、cDNAクロンの従来の蛍光ハイブリダイゼーション（「FISH」）は、1段階における正確な染色体位置を提供するために使用できる。この技術は50から600の長さのgDNAで用いることができる。この技術に関する文献に関しては、Vermaら、Human Chromosomes: A Manual of Basic Techniques、ペルガモンプレス、ニューヨーク（1988）を参照されたい。一度配列を正確な染色体位置にマッピングすると、染色体上の配列の物理学的な位置が遺伝地図データで補正できる。このようなデータは例えば、ジョーンズホプキンス大学、ウェルチメディカルライブラリーを介したオンラインで利用できるV.M

cKusick、Mendelian Inheritance in Manに見出すことができる。同一の染色体領域にマッピングされた遺伝子および疾患との関係は、次いで連鎖分析（物理的に近接した遺伝子の共同遺伝（coinheritance））により同定される。

【0118】次に、cDNAまたはゲノムDNA配列における影響を受けた個体および受けていない個体との差を決定する必要がある。影響を受けた個体の幾つかまたは全てで突然変異が観察されるが、任意の正常の個体では突然変異が観察されない場合、突然変異は疾患の原因物質になり得る。物理的マッピングおよび遺伝的マッピング技術の現在の解析により、疾患に関連する染色体領域に正確に位置するgDNAは50～500の強力原因遺伝子のgDNAであろう。（これは1メガ塩基マッピング解析および20キロ塩基あたり1遺伝子とみなす）。

【0119】ポリペプチド検定

本発明はまたポリペプチドの正常および異常レベルの決定を含む、細胞、組織および体液中のカテプシンKタンパク質のレベルの検出のための定量、診断検定等の診断検定にも関する。本発明の診断法に有用な体液には、例えば滑液、脳脊髄液、尿、血清、歯肉液およびリンパ液等がある。このように、例えば本発明に準じた、正常対照組織試料と比較したカテプシンKの過剰発現を検出する診断検定は、疾患例えば骨粗鬆症、パジェット病、ゴーチャーズ病、CNS炎症、アルツハイマー病、上皮小体亢進症、骨分解、転移性腫瘍、リウマチ性関節症、変形性関節症、歯周病並びに骨インプラントおよび骨プロシーゼスとりわけ歯科インプラントの分解等の存在を検出するために用いることができる。宿主由来の試料中のタンパク質のレベルを測定するために用いることができる検定技術、例えば本発明のカテプシンKタンパク質のイムノアッセイは当業者に周知である。このような検定方法にはラジオイムノアッセイ、競合結合検定、ウェスタンブロット分析およびELISA検定等がある。このうちELISAが頻繁に好ましく用いられる。ELISA検定ではまず、カテプシンKに特異的な抗体、好ましくはモノクローナル抗体を調製する。さらにモノクローナル抗体に結合するリポーター抗体を通常調製する。リポーター抗体は放射性、蛍光または酵素試薬、この例においては西洋ワサビペルオキシダーゼ等の検出可能な試薬に結合する。

【0120】ELISAを実施するために、試料を宿主から取り出し、試料中のタンパク質と結合する固体支持体例えばポリスチレン皿上でインキュベートする。皿上の任意のフリーのタンパク質結合部位をウシ血清アルブミンのような非特異的タンパク質とインキュベートすることにより被覆する。次に、モノクローナル抗体がポリスチレン皿に付着した任意のカテプシンKタンパク質に付着する時間、モノクローナル抗体を皿の中でインキュ

ベートする。非結合モノクローナル抗体は緩衝液で洗い出す。西洋ワサビペルオキシダーゼに連結したリポーター抗体は皿に載せ、リポーター抗体をカテプシンKに結合した任意のモノクローナル抗体に結合させる。非付着リポーター抗体を次いで洗い出す。比色基質等のペルオキシダーゼ活性のための試薬を次いで皿に加える。1次および2次抗体により、カテプシンKに連結した固定化ペルオキシダーゼは呈色した反応産物を産生する。規定の時間内に増加した色の量は試料中に存在するカテプシンKタンパク質の量を示す。一般に定量結果は標準曲線との比較により得られる。競合検定を用いることができ、この検定では、カテプシンK特異的抗体を固体支持体および標識化カテプシンKに付着させ、宿主由来の試料を固体支持体を通して、固体支持体に付着した検出される標識量は、試料中のカテプシンK量に相関する。

【0121】抗体

ポリペプチド、それらの断片もしくはその他の誘導体、またはそれらの類似体、またはそれらを発現する細胞は、それらの抗体を産生する免疫原として用いることができる。これらの抗体は例えばポリクローナルまたはモノクローナル抗体でよい。本発明はまたキメラ、一本鎖およびヒト化抗体、およびF a b断片またはF a b発現ライブラリーの産物等を含む。当業界に周知の種々の方法は、このような抗体および断片の産生に使用できる。本発明の配列に対応するポリペプチドに抗して生じる抗体は、ポリペプチドを好ましくはヒトはでない動物に直接注射するかまたは動物にポリペプチドを投与することにより得ることができる。このようにして得られた抗体は次いでポリペプチド自身に結合する。この方法でポリペプチドの断片のみをコードする配列さえももとの全ポリペプチドに結合する抗体の産生に用いることができる。このような抗体はポリペプチドをポリペプチドを発現する組織から単離するのにもちいることができる。

【0122】モノクローナル抗体の調製のために、連続的なクローンのセルライン培養により産生される抗体を提供する任意の技術を用いることができる。実例としては、ハイブリドーマ法 (Kohler, G. および Milstein, C., Nature, 256:495-497 (1975))、トリオーマ法、ヒトBセルハイブリドーマ法 (Kozborら, Immunology Today, 4:72 (1983)) およびヒトモノクローナル抗体を産生するためのEBV-ハイブリドーマ法 (Coleら, Monoclonal Antibodies and Cancer Therapy, アランR リスインコーポレーティッド、77-96頁 (1985)) 等がある。一本鎖抗体の産生のために記載された技術 (米国特許第4946778号) は本発明の免疫原ポリペプチド産物に対する一本鎖抗体を産生するのに適用できる。また、トランスジェニックマウスまたはその他の生物例えばその他の哺乳動物は、例えば本発明の免疫原ポリペプチド産物に抗するヒト化抗体等の抗体を発現するのに用いることができる。

【0123】従って、とりわけこのような抗体は、突然変異カテプシンKまたは異常カテプシンKレベルに引き起こされるまたはこれらが関与する疾患、例えば骨粗鬆症、歯周病、パジェット病、ゴーチヤーズ病、CNS炎症、アルツハイマー病、上皮小体亢進症、骨分解、転移性腫瘍、並びに骨インプラントおよび骨プロシース (protheses) とりわけ歯科インプラントの分解等の検出および処置に用いることができる。本発明のポリヌクレオチドを用いた免疫化は、カテプシンK特異的免疫応答を生じる周知の反応を用いて実施できる。

【0124】臨床ゲノム学 (Clinical Genomics)

本発明はカテプシンK遺伝子突然変異またはカテプシンK遺伝子発現異常を有するまたはそれらの恐れのある個体の薬物反応性を測定する方法を提供し、またこのような方法を実施する試薬をも提供する。個体は、カテプシンK遺伝子の突然変異またはカテプシンK遺伝子発現により引き起こされるまたはそれらに関与する疾患の処置のために用いられる規定の化合物、とりわけ薬物に対する反応性によりグループ分けできる。このような個体は異なる遺伝子突然変異または遺伝子発現レベル変異型を検出することによりさらにグループ分けできる。このように、特異的遺伝子突然変異および遺伝子発現変異型はある程度個々の化合物に対する反応性と容易に関連付けることができる。本明細書で提供する方法および試薬は、カテプシンK遺伝子突然変異およびカテプシンK遺伝子発現変異型を検出することにより化合物反応性をグループ分けするために用いることができる。化合物反応性による個体のグループ分けのその他の方法は、当業者に周知であり、本発明のポリペプチドおよびポリヌクレオチドの使用に適用できる。本発明はまた、装置と組み合わせて有用な、または組成物として具体化したアルゴリズムを提供し、これはカテプシンKまたはそれらの突然変異もしくは変異型に引き起こされるまたはそれらに関与する疾患の診断に有用である。好ましいアルゴリズムは疾患の階層化および段階化のために提供される。

【0125】カテプシンK結合分子および検定

本発明はまた、本発明のカテプシンKまたはカテプシンK断片に結合する受容体分子のような分子の同定方法をも提供する。カテプシンKに結合するタンパク質例えば受容体タンパク質をコードする遺伝子は当業者に周知の多くの方法、例えばリガンドバニングおよびFACS分類により同定できる。このような方法は多くの実験マニュアル、例えばColiganら、Current Protocols in Immunology、1(2):第5章 (1991) に記載されている。

【0126】例えば発現クローニングはこの目的で用いることができる。この目的のためのポリアデニル化RNAは、カテプシンKに反応する細胞から調製し、cDNAライブラリーはこのRNAから作り、ライブラリーはプールに分割し、プールはカテプシンKに反応しない細胞に個々にトランスフェクトする。トランスフェクトし

た細胞を次に標識化カテブシンKに曝露する。(カテブシンKは放射性ヨウ素化またはプロテインキナーゼ特異的部位の認識部位の封入標準法等の種々周知の技術により標識化できる。)曝露した後、細胞を固定し、カテブシン結合またはカテブシンKに結合する分子を測定する。これらの方法は便宜上ガラススライド上で実施する。

【0127】プールはカテブシンK結合細胞を産生するcDNAを同定する。サブプールは上述のように宿主細胞にトランスフェクトし、スクリーニングしたこれらの陽物質から調製する。サブプール化および再スクリーニング工程を繰り返し、推定される結合分子または基質例えば細胞マトリックス、骨マトリックスまたは受容体分子をコードする1またはそれ以上の単一のクローン、および同様のものはどれも単離できる。また別に、標識化リガンドは細胞抽出物例えばリガンドが結合する分子例えば受容体分子を発現する細胞から調製した膜または膜抽出物に連結した光親和物にできる。交差架橋物質はポリアクリルアミドゲル電気泳動(「PAGE」)により解析し、X線フィルムに曝露する。リガンド受容体を含む標識化複合体は切り取り、ペプチド断片を分解し、タンパク質のマイクロシーケンシングに供することができる。マイクロシーケンシングにより得られたアミノ酸配列は、推定される受容体分子をコードする遺伝子を同定するcDNAライブラリーをスクリーニングする、独特なまたは同義性オリゴヌクレオチドプローブの設計に用いることができる。本発明のポリペプチドはまた、細胞中のまたは細胞不含調製物中の、カテブシンK結合分子、例えば受容体分子のカテブシンK結合能力を評価するために用いることができる。

【0128】アゴニストおよびアンタゴニスト-検定および分子

本発明はまた細胞中または上のカテブシンKの作用、例えばカテブシンK結合分子例えば受容体および酵素基質分子とカテブシンKの相互作用を増強または阻止する化合物を同定するための、化合物のスクリーニング方法をも提供する。アゴニストはカテブシンKの天然の生物学的機能を増強する化合物であり、一方アンタゴニストはこのような機能を低下または排除する化合物である。例えば、細胞の区画例えば膜、空胞、封入物またはそれらの任意の調製物例えば膜調製物は、カテブシンKに結合する分子、例えばカテブシンKにより調節される信号化または制御経路の分子を発現する細胞から調製できる。その調製物を、カテブシンKアゴニストまたはアンタゴニストに成り得る候補分子の存在下または不在下で、標識化カテブシンKと共にインキュベートする。候補分子が基質等の結合分子に結合する能力は、標識化リガンドの結合の低下の影響を受ける。結合しても影響を及ぼさない分子、すなわちカテブシンK結合分子の結合にカテブシンKの影響を誘起しない分子は、最も良好なアンタ

ゴニストに成るであろう。結合性が良好で、カテブシンKと同一のまたは非常に関連した効果を引き出す分子は、アゴニストである。

【0129】有効なアゴニストおよびアンタゴニストのカテブシンK様効果は、例えば候補分子の細胞または適当な細胞調製物との相互作用に続く2次メッセンジャーシステムの活性を決定し、効果をカテブシンKの効果とまたはカテブシンKと同一の効果を引き出す分子の効果と比較することにより測定できる。この点に関して有用に成り得る2次メッセンジャーシステムには、AMPグアニレートサイクラーゼ、イオンチャネル、ホスホノシタイド加水分解2次メッセンジャーシステム、またはカテブシンKの強力なアゴニストおよびアンタゴニスト、またはその基質の結合を信号化する化合物等があるが、これらに限定するものではない。

【0130】カテブシンKアンタゴニストの検定の他の例は競合検定であり、競合阻害検定に適当な条件下で、カテブシンKおよび強力なアンタゴニストを酵素基質または基質類似体に結合させる。カテブシンKを例えば放射活性により標識化し、受容体分子に結合するカテブシンK分子の数を正確に決定し、有効なアンタゴニストの効果を評価できる。

【0131】有効なアンタゴニストには、本発明のポリペプチドに結合し、それによりその活性を阻害または消失させる小さい有機分子、ペプチド、ポリペプチドおよび抗体等がある。有効なアンタゴニストにはまた、結合分子例えば受容体分子の同一の部位で結合する、カテブシンK誘起活性を誘起せず、それによりカテブシンKを結合から排除することによりカテブシンKの作用を防御する小さい有機分子、ペプチド、ポリペプチド、例えば非常に相関したタンパク質または抗体等がある。有効なアンタゴニストには、ポリペプチドの結合部位に結合し、結合部位を占領し、それにより細胞結合分子例えば受容体分子への結合を防御し、正常の生物学的活性を防御する、小さい分子等がある。小さい分子の例としては、小さい有機分子、ペプチドまたはペプチド様分子などがあるが、これらに限定するものではない。

【0132】その他の有効なアンタゴニストにはアンチセンス分子等がある。アンチセンス技術はアンチセンスDNAまたはRNAによる、または三重らせん形成による制御遺伝子発現に用いることができる。アンチセンス技術については、例えばOkano, J., *Neurochem.* 56:560 (1991); *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRCプレス、ボッカラートン、フロリダ州(1988)に記載されている。三重らせん形成については例えばLeeら、*Nucleic Acids Research* 6:3073 (1979); Cooneyら、*Science* 241:456 (1988); およびDervanら、*Science* 251:1360 (1991)に記載されている。この方法は、ポリヌクレオチドの相補的DNAまたはRNAへの結合に基づいている。例えば本発

明の成熟ポリペプチドをコードするポリヌクレオチドの5'コード部分は、約10～40塩基対の長さのアンチセンスRNAオリゴヌクレオチドを設計するのに用いることができる。DNAオリゴヌクレオチドは、カタブシンKの転写および産生を防御する転写に係わる遺伝子領域に相補的になるように設計される。アンチセンスRNAオリゴヌクレオチドはインビボでmRNAにハイブリダイズし、mRNA分子のカタブシンKポリペプチドへの翻訳を阻止する。上述のオリゴヌクレオチドはまた、アンチセンスRNAまたはDNAがインビボで発現し、カタブシンKの産生を阻害するように細胞に運ぶこともできる。

【0133】このアンタゴニストは、例えば本明細書の以下に記載するような、医薬的に許容され得る担体と共に組成物として用いることができる。このアンタゴニストは、例えば突然変異カタブシンKまたは異常カタブシンKレベルにより引き起こされるまたはこれらに関与する疾患、例えば骨粗鬆症、パジェット病、ゴーチアーズ病、CNS炎症、アルツハイマー病、上皮小体亢進症、骨分解、転移性腫瘍、リウマチ性関節症、変形性関節症、歯周病および骨インプラントおよび骨プロシーゼスとりわけ歯科インプラントの分解等の処置に用いることができる。

【0134】組成物

本発明はまた上述のポリヌクレオチドもしくはポリペプチド、またはアゴニストもしくはアンタゴニストを含む組成物にも関する。従って、本発明のポリペプチドは非無菌もしくは無菌担体と、または細胞、組織もしくは生物に用いられる担体、例えば対象への投与に適した医薬的担体と組み合わせて用いることができる。このような組成物には例えば媒体添加物または本発明のポリペプチドの治療の有効量、および医薬的に許容できる担体または賦形剤を含む。このような担体には生理食塩水、デキストロース、水、グリセロール、エタノールおよびそれらの組み合わせ等であり、これらに限定するものではない。製剤は投与法に適した物にすべきである。

【0135】キット

本発明はさらに、本発明の前述の組成物成分を1またはそれ以上を充填した1またはそれ以上の容器を含む医薬的バックおよびキットにも関する。このような（複数の）容器と一緒に、医薬または生物学的産物の製造、使用または販売を統制する政府機関により処方された形態の注意書であって、ヒトへの投与用の製品の製造、使用または販売に関する該政府機関の承認を示す注意書が付されることがある。

【0136】投与

本発明のポリペプチドおよびその他の化合物は、単独でまたは治療用化合物等のその他の化合物と組み合わせて用いることができる。医薬組成物は任意の有効な、利便的な方法例えば、とりわけ局所、経口、経膈、静脈内、

腹腔腔内、筋肉内、皮下、鼻腔内、関節内または皮膚内の経路で投与できる。医薬組成物は、通常具体的な適応症または複数の適応症の処置または予防に有効量投与する。一般に、この組成物は少なくとも約10mg/体重の量で投与する。好ましくは、たいの場合、投与量は1日あたり約10mg/kgから1mg/kg体重である。至適投与量は、適応、その重篤度、投与経路、合併症の状態等を考慮に入れて、各々の処置様式および適応のための標準法により決定されることは理解されよう。

【0137】遺伝子治療

カタブシンKポリヌクレオチド、ポリペプチド、ポリペプチドであるアゴニストおよびアンタゴニストは、本発明に準じて、インビボでこのようなポリペプチドを発現することにより、しばしば「遺伝子治療」と称せられる治療様式に用いることができる。このように、例えば患者の細胞をエクソビボでポリペプチドをコードするDNAまたはRNAのようなポリヌクレオチドで操作でき、次いで操作した細胞をポリペプチドで処置すべき患者に提供できる。例えば、本発明のポリペプチドをコードするRNAを含有するレトロウイルスプラスミドベクターを使用することにより、細胞をエクソビボで操作できる。このような方法は当業界で周知であり、本発明における使用は、本明細書の教示より明白である。

【0138】またポリヌクレオチド例えば構築されたりボザイムで、周知の方法を用いて患者の細胞を操作し、カタブシンKの発現を阻害することができる。その他の構築物はまた、周知の方法を用いて、患者の細胞に操作し、カタブシンKのアンチセンス配列を含有させることができる。このようなアンチセンス構築物は患者のカタブシンK発現を阻害する。

【0139】同様に、細胞を当業界に周知の方法によりインビボで操作し、インビボでポリペプチドを発現させることができる。例えば本発明のポリヌクレオチドは上述のように複製不能レトロウイルスベクター中に発現するように操作できる。レトロウイルス発現構築物を、次いで単離し、本発明のポリペプチドをコードするRNAを含有するレトロウイルスプラスミドベクターで形質導入されたパッケージング細胞に導入し、パッケージング細胞が目的遺伝子を含有する感染性のウイルス粒子を産生するようにすることができる。これらの産生細胞は患者に投与し、インビボで細胞操作し、インビボでポリペプチドを発現することができる。このような方法で本発明のポリペプチドを患者に投与するためのこれらのまたはその他の方法は、本明細書の教示から当業者に明らかであるはずである。

【0140】上述のレトロウイルスプラスミドベクターは、モロニーネズミ白血病ウイルス、脾臓壊死ウイルス、レトロウイルス例えばラウス肉腫ウイルス、ハーヴェイ肉腫ウイルス、鳥白血症ウイルス、テナガザル白血病ウイルス、ヒト免疫不全ウイルス、アデノウイルス、

骨髄増殖性肉腫ウイルスおよび哺乳動物腫瘍ウイルス等に由来するが、これらに限定するものではない。一つの態様では、レトロウイルスプラスミドベクターはモロニーネズミ白血病ウイルスに由来する。このようなベクターは、ポリペプチド発現のためのプロモーターを1またはそれ以上含む。用いることができる適当なプロモーターには、カテプシンKプロモーター、レトロウイルスLTR、SV40プロモーター、およびヒトサイトメガロウイルス (CMV) プロモーター (Millerら、Biotechniques 7:980-990 (1989) に記載されている)、またはその他の任意のプロモーター (例えばヒストン、RNAポリメラーゼ I II およびアルファアクチンプロモーター等の真核細胞プロモーターのような細胞性プロモーターなどがあるが、これらに限定するものではない) 等があるが、それらに限定するものではない。用いることができるその他のウイルス性プロモーターには、アデノウイルスプロモーター、チミジンキナーゼ (TK) プロモーター、およびB19パルボウイルスプロモーター等があるが、これらに限定するものではない。適当なプロモーターの選別は、本明細書の教示より、当業者に明白であらう。

【0141】本発明のポリペプチドをコードする核酸配列は適当なプロモーターの制御下に置かれる。用いることができる適当なプロモーターには、アデノウイルスプロモーター、例えばアデノウイルス主要後期プロモーター；または異型性プロモーター例えばサイトメガロウイルス (CMV) プロモーター；ラウス肉腫ウイルス (RSV) プロモーター；誘導プロモーター例えばMMTプロモーター、メタロチオネインプロモーター；熱ショックプロモーター；アルブミンプロモーター；アポA I プロモーター；ヒトグロブリンプロモーター；ウイルス性チミジンキナーゼプロモーター例えば単純ヘルペスチミジンキナーゼプロモーター；レトロウイルスLTR (上述の改変レトロウイルスLTRを含む)；アルファアクチンプロモーター；ヒト成長ホルモンプロモーター等があるが、これらに限定するものではない。プロモーターはまたポリペプチドをコードする遺伝子を制御する天然プロモーターでもよい。

【0142】レトロウイルスプラスミドベクターはパッケージングセルラインに形質導入し、産生セルラインを形成するのに用いることができる。トランスフェクトされ得るパッケージング細胞の例としては、PE501、PA317、Y-2、Y-AM、PA12、T-19-14X、VT-19-17-H2、YCRE、YCRIP、GP+E-86、GP⁺envAm12およびDANセルライン (Miller, A., Human Gene Therapy 1:5-14 (1990) に記載されている) 等があるが、これらに限定するものではない。ベクターは当業界で周知の任意の方法により、パッケージング細胞に形質導入できる。このような方法には、電気泳動、リボソームの使用、およびC

aPO₄沈殿などがあるが、これらに限定するものではない。また別に、レトロウイルスプラスミドベクターはリボソームでカプセル化するか、または脂質と結合させ、次いで宿主に投与することができる。

【0143】産生セルラインは感染性レトロウイルスベクター粒子を生じ、これはポリペプチドをコードする (複数の) 核酸配列を含む。このようなレトロウイルスベクター粒子は次いでインビトロまたはインビボのどちらかで真核細胞に形質導入するのに用いることができる。形質導入した真核細胞は、ポリペプチドをコードする (複数の) 核酸配列を発現する。形質導入された真核細胞には、胚幹細胞、胚癌細胞およびヘモポイエシス幹細胞、肝細胞、繊維芽細胞、筋芽細胞、ケラチノサイト、内皮細胞および気管支上皮細胞等があるが、これらに限定するものではない。

【0144】

【実施例】本発明はまた以下の実施例によりさらに記載する。この実施例は単に本発明の説明のために具体的な態様の参考に提供するものである。これらの例示は本発明の特定の具体的な態様を説明するものであるが、開示する発明の範囲の限定または制限を表すものではない。本明細書で用いる特定の用語は、前述の定義で説明する。本明細書で用いるヌクレオチド配列におけるNは、未知の一個のヌクレオチドまたは複数のヌクレオチドを意味する。

【0145】全ての実施例は、他に詳細に記載するもの以外は、標準的な技術を用いて実施した、または実施でき、これは当業者に周知で通常的である。以下の実施例の通常的な分子生物学的技術は、標準的な実験マニュアル、例えば本明細書で "Sambrook" と称する、Sambrookら、Molecular Cloning: A Laboratory Manual、第2編；コールドスプリングハーバーラボラトリープレス、コールドスプリングハーバー、ニューヨーク州 (1989) に記載されるように実施できる。以下の実施例で示す全ての部分または量は、特に限定しない限り、重量で示す。

【0146】以下の実施例における断片のサイズ分離について、他に記載しない場合、Sambrookおよび多くのその他の参考文献例えばGoeddelら、Nucleic Acids Res. 8:4057 (1980) に記載される、寒天およびポリアクリルアミドゲル電気泳動 ("PAGE") の標準技法を用いて実施した。他に記載しない場合、ライゲーションは標準的な緩衝液、インキュベーション温度および時間、ライゲートすべきDNA断片の約等モル濃度の量で、DNA 0.5 μg あたり約10単位のT4 DNAリガーゼ ("リガーゼ") を用いて行った。

【0147】実施例1

ヒトカテプシンKゲノムクローンの単離およびシーケンシング

以下の方法に準じて、米国特許第5501969号に開

示される cDNA を、gDNA ライブラリー (クローニング) から gDNA クローンを単離するのに用いた。近接するエキソン (7 エキシソンの 6) のプライマーを調製した。これらのプライマーの配列は図 10 および 11 で下線を付している。PCR は当業界で周知の標準法を用いて実施した。増幅断片は TA ベクター (クローニング) にクローン化し、正および逆方向シーケンシングプライマーを用いて当業界に周知の確立された方法により、自動シーケンサー (応用型バイオシステムモデル 373) でクローンを配列決定した。すべての内在イントロンの配列を得た。5' および 3' 末端イントロン配列は以下のようにして得た。最初のイントロンのための配列 (図 10 および 11 の下線を付したプライマーを参照) を得るために、5' 末端プライマーを設計し、これらのプライマーを用いて 2 P1 クローンを得た (ゲノムシステムズインコーポレーティッド)。両方のクローンは全長である。内部のイントロン-エキソン境界接合部の配列を確認するために PCR を使用した (実施例 2 参照)。P1 クローンの 5' 末端の配列に由来するプライマーは、当業界周知の通常的な方法により、各配列決定段階での新規のプライマーを用いて、段階的な方法で「ウォーク (walk)」し、クローンに沿った配列決定をするために使用することができる。P1 クローンの精製は、実施例 1 (d) に示す様に実施した。ウォーキングおよび配列決定は両方向で実施し、カテプシン K gDNA 配列を確認した。校正 Taq ポリメラーゼ (PCR ウルチマ、ペルキンエルマー) を用いて再び PCR を実施した。転写開始部位は、5' RACE キット (ギブコ BRL) およびそこから供給されたプロトコルを用いて得た。この部位はまた RNアーゼ保護検定キット (ハイブスピード、RPA アンピオン) を用いても確認した。実施例 1 (a) - (d) はカテプシン K のクローニングおよびシーケンシングに関してさらに具体的に記載している。

【0148】 (a) イントロン-エキソン境界の DN A シーケンシング

イントロン-エキソン境界

イントロン 1

エキソン 1 から 2 の PCR に設計され得るプライマーの 5' UTR 配列を決定する 5' RACE (ギブコ BRL) を利用して、イントロン 1 を同定した。(イントロン 1 は ATG の前に出発するので、PCR は可能な cDNA 配列に基づいて容易には用いることができない) イントロン 1 はヒトゲノム DNA (クローニング) で PCR により増幅し、PCR II ベクターにクローン化し、実施例 1 に記載されるように配列決定した。

イントロン 2

イントロン 2 はヒトゲノム DNA で、エキソン 2 からエキソン 3 に設計したプライマーから、PCR により同定した。PCR 産物は標準法を用いてクローン化および配

列決定した。

イントロン 3

イントロン 3 はヒトゲノム DNA で、エキソン 3 からエキソン 4 に設計したプライマーから、PCR により同定した。PCR 産物は標準法を用いてクローン化および配列決定した。

イントロン 4

イントロン 4 はヒトゲノム DNA で、エキソン 4 からエキソン 5 に設計したプライマーから、PCR により同定した。PCR 産物は標準法を用いてクローン化および配列決定した。

イントロン 5 & 6

イントロン 5 および 6 はヒトゲノム DNA で、エキソン 5 からエキソン 7 に設計したプライマーから、PCR により同定した。PCR 産物は標準法を用いてクローン化および配列決定し、両イントロンの存在を確認した。

イントロン 7

イントロン 7 はヒトゲノム DNA で、エキソン 7 からエキソン 8 に設計したプライマーから、PCR により同定した。PCR 産物は標準法を用いてクローン化および配列決定した。ヒトゲノム DNA で PCR により同定した全てのイントロンは、P1 クローン A (以下の (b) 参照) クローン (ゲノムシステムズインコーポレーティッド) の同一の領域の PCR により確認した。

【0149】 (b) 5' および 3' 非翻訳領域 (UTR) の DNA シーケンシング

5' および 3' 非翻訳領域は単一の P1 クローン (ゲノムシステムズインコーポレーティッド) から単離した。この P1 クローンは本明細書では "P1 クローン A" と同定した。標準法を用いて確認した cDNA 配列に由来する遺伝子特異的プライマーで、P1 クローンの直接的に配列をウォーキングアップおよびダウンすることにより配列決定した。これらの領域を次に PCR によりクローン化し、標準法を用いて配列分析により確認した。製造者指示書に準拠して、校正 Taq ポリメラーゼウルチマを用いて、PCR により 5' UTR をさらに増幅し、クローン化し配列の不明瞭性を排除した。5' および 3' UTR はさらに標準法を用いて、ヒトゲノム DNA 上で PCR により確認した。

【0150】 (c) mRNA キャップ部位の DNA シーケンシングおよびエキソン 1 の大きさ

mRNA キャップ部位は、5' RACE シーケンシングに基づき、出発コドンの約 48 塩基対上方にあると決定された。リボヌクレアーゼ保護検定は転写からの出発部位が ATG (出発コドン) の約 48 塩基対上方にあることを示唆する、約 48 塩基対の大きさの保護断片を確認した。推定される転写因子は、データベース転写因子配列情報で配列分析して同定し、これらは図 23 - 24 (S) [配列番号: 2] に示す。1.1 キロ塩基の 5' UTR 断片は pCAT 発現ベクターにクローン化

し、さらにプロモーター配列領域を分析した。

【0151】(d) P1 DNA調製

カナマイシンLBプレート上でP1クローンAコロニーを筋付けした。単一のコロニーを採り、 $25\mu\text{g}/\text{ml}$ のカナマイシンと共に 20ml 中でO/N成長させた。O/N培養物 16ml を培地 500ml (カナマイシン $25\mu\text{g}/\text{ml}$)に接種し、10時間成長させた。細胞は遠心によりペレット化し、キアゲンP1溶液 10ml に再懸濁した。キアゲンP2溶液 10ml を加え、室温で5分間インキュベートした。キアゲンP3溶液 10ml を加え、混合物を15分間氷上で放置した。試料を 10000g で15分間回転させた。上澄を除去し、フェノールで抽出した。上澄を次にクロロホルムで再抽出した。 NaOAc pH5.2およびイソプロパノール 1.1 容量を加えてDNAを沈殿させた。 10000g で15分間遠心して、DNAをペレット化し、70%エタノールで洗浄した。シーケンシング用にDNAを洗浄するために、DNA $250\mu\text{l}$ (約 $50\mu\text{g}$)を 1.5M の NaCl 中30%PEG65 μl に加えた。 3M NaCl $8.5\mu\text{l}$ を加え、混合物を氷上で30分間インキュベートした。試料を 12000g で10分間回転させた。上澄を捨て、ペレットを $200\mu\text{l}$ の蒸留水に溶解した。DNAを次にクロロホルムで抽出し、旋回し 10000g で1分間回転させた。水層を除去し、DNAを $40\mu\text{l}$ の NaOAc pH5.2および 1ml のエタノールで沈殿させた。試料を 12000g で30分間回転させた。DNAを70%エタノール 1ml で洗浄し、蒸留水に再懸濁した。シーケンシングの前にDNAを 2M NaOH 0.1 容量および 2mM EDTAで変性し、 37°C で30分間インキュベートした。混合物を 3M NaOAc pH5.2、 0.1 容量で中和し、エタノール 2.5 容量で沈殿させた。変性DNAを蒸留水中 $1\mu\text{g}/\mu\text{l}$ 、 $6\mu\text{g}/\mu\text{l}$ の濃度で再懸濁し、TaqFSを用いる各々のシーケンシング反応(ABI)で使用した。

【0152】実施例2

カテプシンKの染色体マッピング

精製P1 DNAを具体的な染色体にマッピングするために、FISH分析(ゲノムシステムズインコーポレーティッド)に用いた。2PCR体細胞ハイブリッドパネルの使用により、先に得られた結果は、染色体1へマッピングした遺伝子を示した。FISH分析により、染色体1へのマッピングおよびさらに $1q21$ にマッピングした遺伝子を確認した。これはカテプシンSとして知られている座と同一の座である。

【0153】実施例3

COS細胞におけるカテプシンKの発現

CAT検定

CATリポーター遺伝子の上方に 1100 塩基対の推定CatKプロモーターを含有するpCAT-CatKは、DEAE-デキストラン法によりCOS細胞にトラ

ンスフェクトした。トランスフェクションは 100mm 皿中COS細胞に施し、 $5\mu\text{g}$ のDNAを用いた。対照として、プロモーターまたはエンハンサーを含まないpCATベシックおよびSV40プロモーターおよびエンハンサーを含むpCATコントロールをまた別々に移した。トランスフェクションの72時間後、凍結融解により抽出物を作り、抽出タンパク質の等量を1時間および一晩の両方のCAT検定に用いた。非トランスフェクトCOS細胞では活性は検出されなかった。pCAT-CatKは、非トランスフェクト細胞からのバックグラウンドを減じた後、pCATベシックに相対的なCAT発現の $1.4\sim 1.6$ 倍の増加を示した。種々のインデューサーの存在下高レベルの活性化を得ることが可能であるので、外因性 $1, 25$ ジヒドロキシビタミンD3の添加によりCatKプロモーターの活性化が起こると考えられる。ビタミンDは他の研究によりオステオカルシン、オステオポンチン、カルシトニンおよびP450プロモーターの転写を、ビタミンD受容体およびこれらの種々プロモーターに見出される(複数の)ビタミンD反応要素との相互作用により活性化することが示されている。ビタミンDがこれらのプロモーターの活性化移動する能力は骨形成および再吸収の制御において役割を果たしていると考えられる。エストロゲン反応性を評価するために、類似した実験を行い、これもまた骨形成および再吸収の制御において役割を果たしていると考えられる。

【0154】実施例4

ヒトカテプシンKの遺伝子治療的発現

皮膚生検により対象から繊維芽細胞を得た。得られた組織を組織培養培地に載せ、小片に分割する。組織の小さい固まりを組織培養フラスコの湿った表面上に載せ、約10片を各フラスコに載せる。フラスコを上下逆に置き、きっちり締めて一晩室温で放置する。室温で24時間後、組織の固まりはフラスコの底に付着して残るようにしてフラスコを逆さにし、新鮮な培地を加える(例えば10%FBS、ペニシリンおよびストレプトマイシンを加えたHam's F12培地)。組織を次いで 37°C で約1週間インキュベートする。この時、新鮮培地を加え、続いて数日毎に変える。更なる2週間の培養の後、繊維芽細胞の単層が生じる。その単層をトリプシン処理し、大きなフラスコに剥ぎおとす。遺伝子治療のためのベクターは、発現させる断片をクローニングするために制限酵素で消化した。消化したベクターを仔ウシ腸管ホスファターゼで処理し、自身のライゲーションを防御する。脱リン酸化した直線ベクターは寒天ゲル上で分画化し、精製する。

【0155】活性カテプシンKを発現する能力のあるカテプシンKg DNAを単離する。好ましい構築物は、細胞型特異性遺伝子発現のためにカテプシンKプロモーターを用いる。断片の末端は、必要な場合、ベクターにク

ローニングするために改変する。例えば、5' 張り出しはDNAポリメラーゼで処理して平板な末端にできる。3' 張り出しはS1ヌクレアーゼを用いて除去できる。リンカーはT4 DNAリガーゼで平板末端にライゲートできる。モロニーネズミ白血病ウイルスの直線的な骨格およびカテプシンK断片の等容量を混合し、T4 DNAリガーゼを用いて結合した。ライゲーション混合物は、大腸菌を形質転換するために用い、細菌を次いでカナマイシン含有寒天上に載せる。カナマイシン表現型および制限分析でベクターが遺伝子を適切に挿入したことを確認する。10%仔ウシ血清(CS)、ペニシリンおよびストレプトマイシン含有デュルベッコ改変イーグル培地(DMEM)中、全面成長するまでパッケージング細胞を組織培養で成長させる。カテプシンK遺伝子を含有するベクターを標準法によりパッケージング細胞に導入する。カテプシンK遺伝子を含有する感染性ウイルス粒子をパッケージング細胞から回収し、これを産生細胞と称する。

【0156】新鮮な培地を産生細胞に加え、適当なインキュベーション時間の後、全面成長した産生細胞のプレートから培地を収穫する。感染ウイルス粒子を含有する培地をミリポアーフィルターを通してろ過し、剥がれた産生細胞を除去する。ろ過した培地を次に繊維芽細胞に感染させるために使用する。培地を繊維芽細胞の副全面成長プレート(sub-confluent plate)から除去し、すぐにろ過した培地と入れ替える。ポリブレン(アルドリッヒ)は培地中に含まれ、形質導入を促進できる。適当なインキュベーションの後、培地を除去し、新鮮な培地と入れ替える。ウイルスの力価が高い場合、実質的に全ての繊維芽細胞が感染し、選別の必要はない。力価が低い場合、ネオまたはヒズのような選別マーカーを有するレトロウイルスベクターを用いる必要があり、発展させるための形質導入細胞を選別する。次いで操作した繊維芽細胞を、単独またはマイクロキャリアビーズ例えばサイトデックス3ビーズに全面成長させた後、ヒトまたは動物例えばラットおよびマウスに注入できる。注入した繊維芽細胞はカテプシンKを産生し、タンパク質の生物学的作用を宿主に伝達する。

【0157】本発明は、前述の説明および実施例に特に記載した以外も、実行できることは明らかである。上述の教示に鑑みて、本発明の多くの改変および変形が可能であり、従ってそれらも添付の請求の範囲の範疇である。

【0158】

【配列表】

(1) 一般的情報:

(i) 出願人: デヴーク、クリスティーヌ、ドレイク、フレッド、ゴーウェン、マキシム、ロッド、ジュリー、ヘイスティングス、グレッグ、アダムス、マーク、フレイザー、クレア、リー、ノーマン、カークネス、イーウ

エン、ブレイク、ジュディス、フィッツジェラルド、リサ

(ii) 発明の名称: カテプシンK遺伝子

(iii) 配列の数: 32

(iv) 連絡先:

(A) 名称: スミスクライン・ピーチャム・コーボレイション

(B) 通り名: スウェードランド・ロード709番

(C) 都市名: キング・オブ・プルシア

(D) 州名: ペンシルベニア州

(E) 国名: アメリカ合衆国

(F) 郵便番号: 19406-2799

(v) コンピューター読取り可能な形態:

(A) 媒体形態: ディスケット

(B) コンピューター: IBMコンパチブル

(C) オペレーティングシステム: DOS

(D) ソフトウェア: FastSEQ Version 1.5

(vi) 現出願データ:

(A) 出願番号:

(B) 出願日: 1997年5月7日

(C) 分類:

(vii) 先の出願データ

(A) 出願番号: 60/019,942

(B) 出願日: 1996年6月14日

(vii) 先の出願データ

(A) 出願番号: 60/020,273

(B) 出願日: 1996年6月17日

(vii) 先の出願データ

(A) 出願番号: 60/026,273

(B) 出願日: 1996年8月26日

(viii) 代理人等の情報:

(A) 氏名: ハン, ウィリアム・テイ

(B) 登録番号: 34,344

(C) 代理人等における処理番号: ATG50006-2

(ix) テレコミュニケーションの情報:

(A) 電話番号: 610-270-5219

(B) テレファックス番号: 610-270-5090

(C) テレックス:

【0159】(2) 配列番号: 1の情報:

(i) 配列の特徴:

(A) 配列の長さ: 13674塩基対

(B) 配列の型: 核酸

(C) 鎖の数: 一本鎖

(D) トポロジー: 直鎖状

(ii) 分子の型: ゲノムDNA

(iii) 仮定: なし

(iv) アンチセンス: なし

(v) 断片の型:

(vi) 起源:

(xi) 配列の記載：配列番号：1：

GCTTTGGCTC CCAAAGGCCT GGGATTACAG GCGTGAACCA CTGCGCCTAG CCTGTTAGCA 60
GCTCTTAAAA TCCAGAGGCA TAAGCCTGTA TTTTGGAGG TTTATGCATG GAATCCAGCT 120
AGAAACTGAG TCTATTACAG ATCCCATTTA TTATCCTTTC TATTCCAAGA AGCCTTTTTT 180
TCTCCTTCCC CACATCTGTT TATGGAAGAA AATGAAGTTT GGGGTGTGGT TTGAGGAATC 240
AGCTAGATTC TTATGATCTG TCACATGCTT GGATGTTGGG GAAGCATTTG GAGAAGCTCA 300
TGTGACTTGT CCTAGATTGG GGATTTTAAT TGAGACAGAT GATGTTTATC GGGCATCCCA 360
CCACCTGAGA GTTTTAGCAA CAGAGTCACA TGTGAGTCCA TCAGAACTTA CGGCATTGAT 420
TCAAGTGCTG TCATAAATAA CCAGGACTGC TGTTTTGGT TACTTTTAAA GACAGTTTCA 480
TCTGGACTTT CTGGGCATAT CCTCCTTCAG CAAAACCACA TTAGGCTGGG AAAACTATTC 540
TGCTGGAAG TAATGACAAC TTGCAACCAA CAAGCTTATA AAAATACAAA GAATTCTGGA 600
GCCTATGGCT TCCATTACAT TATTCCTTTA TAGCCTTTTA TGTTCAATAC CGCATCCCAG 660
AGGTGAGAGT CAGACACAAA TATGAAAATA GGTTCATATG TTGGAGAGGT AAATCCTAAC 720
AGGAAAGGGG TAGGAAAAGA TATAATCCCC CAATATTAAA ATAAAGATAT TGAAGAAGAA 780
GGATGGGAGA GACTAGGGCT GTGTCCTTCC TTTTACTCAC CAAAAGAGAA AGTAAGCTCC 840
TATTTGAGTC AATAGATATT GAGGTCTTGT TATTTGCCAC CAAAGACAGT CTTGTGAGAC 900
TAAATAGCTA GTAATTCCTT ACCCTGGCAC ACATGCTGCA TACACACAGA AACACTGCAA 960
ATCCACTGCC TCCTTCCCTC CTCCTACCC TTCCTTCTCT CAGCATTTCT ATCCCGCCT 1020
CCTCCTCTTA CCCAAATTTT CCAGCCGATC ACTGGAGCTG ACTTCCGCAA TCCCGATGGA 1080
ATAAATCTAG CACCCTGAT GGTGTGCCCA CACTTTGCTG CCGAAACGAA GCCAGACAAC 1140
AGATTTCAT CAGCAGGTAA CGTTTGCAAC TTCCTAGATC TTTTAGCTTT TCATTCCTGT 1200
CAATTCTCTG AGTATTAGGG ATGTAGTGAC TTGAGGATCA CAATAAACTT TTAGCCTCTG 1260
CAGATGAAAA CAGAGATGCA CTTCTTAGGT CATTCCCTGG CTAAATAAAA TCTGCCTGGA 1320
AATCTGTAGA ATTCCTTGTA TGATTATAT ATATACATAC ATGATTGTGA GTAAAAGCAA 1380
AGTATATAGG GAATCATTTT CCCATCCTTC AAGAGTGGCC TTTCTGCAGT GTTTTCTACT 1440
TTGGCCAACA AGGATCAAAA CGGTTAACTC CTTAGTGAGG AGGAGGAGAG TGGTATGGGG 1500
AGGTAGTAGC TCAGTGCTTC CTGTTCACTG AGACATCTCA AAGCCCTTAA CACTCTAGTT 1560
TTTAAATGTC CTA CTGAGACA TTTTGCCAGT TTGCAAAATT ACATGTAAAT GGAATAAG 1620
CAATTGTGTA AGCCATATGT CATGCTGCAG GCTGCAAAAT GTTCTTAAAA TGGAGGATTT 1680
GTAATTAAGA AAGCCAATGC AAGAAATGAG TGAAGCTAAC TAGAGTAAAC TTATGAAAAG 1740
CTGTGAATTT CATCATATA GAACATTGCT TTTGAGTCTG AACATTCTTC TAACAAACCT 1800
TGGATCTGAG GCTTCTTGTC CTTTGCGGCA GCCACAGTGG GTTTTTGTG TTAGGGGAAA 1860
ATAAAAAACC TTGCCCGCAG CATCTGGTTA AGATTAGGGC AGTTTCCTGC CTAAGGAGGG 1920
AAGGGAGAGA AAAAGGAAGA AGAAATGCAT AAGGAGAATG AGGAGATATA CAATGTCTCA 1980
GAAAACAGGA AACATTGTCC TATTTTCCCT TGTCCTCTTC TGACAAGATC TGGGAAAGTA 2040
CCAGAATTTA GGCACGAAAG AGAAGAAGC CTGCAAGAAA TGATCAGGAA GCAAACTTA 2100
GACGGAAATC TCTCCTTGT GTATTCTGAA CCCCCTACC ACCTTGCTAT TTGCTGTCT 2160
CCAAGCCTGC TAGGGACCCT GGAGGAAACG CACTGAGCCC ATTCTGATTG TCCAGTTTCT 2220
ATCCCCCATT TCTGGTTGTG TACGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG 2280
TGTGTGTGAG AGAGAGAGAG ACAGAGAGAG AAACAGAGAG AGTGTGTGTT GCCTAAATCT 2340
CCCGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAAAAGAGA GAAATGGCTA 2400
AATCCCCCTA GATCAAAGTC CTGGAACCA GATGTACCAG CATCCTATCT AAACACAGGC 2460
CCCTCCTGAC TATCATTGTT TTATCACCCT TTTTCCGTCT ACCTTTCTCT TCCTCATAAA 2520
GCCTAGTTTT CCTCTGTTT CCTGCCAAAT GGAAGAGTTT TCCCTAACTA CATTCTCTG 2580
CAGGATGTGG GGGCTCAAGG TTCTGCTGCT ACCTGTGGTG AGCTTTGCTC TGTACCTGA 2640
GGAGATACTG GACACCCACT GGGAGCTATG GAAGAAGACC CACAGGAAGC AATATAACAA 2700
CAAGGTGCCT GGGTCTCTGG AGGGGGCATG GCAGGAAGGC TGAGACCTGA GCTCTCTCAT 2760
CTTAGCTTCC AGACTCCCTT CTTCAATCCA AATGCTTTAT TCCAAGCAAA TCAGTCCCTC 2820
TTCCCTAACT CATGTTAACA TACGGTTTTC ATTCCTATGC TTCAATCATC CTCTGTCAA 2880
ACTTGATTC CTTCCCTTTG GTTTTATAAG TGTGTAACAT TCCTCTTTTG GGAAGAGTCC 2940

CAAGATTAAT GCTGTTAATC CATAAGCAAT TTTTCTGTCT CTCCAGAGCT TGTGTGGTGG 3000
 TTTACATATT ATCTCTCTTC TTGCAGGCTC TTAATTCAT GGTAGTTCC CCAACTAAAC 3060
 TGTAAACTTT TATGATTGTG AGTTTCCTTT ATTCTCCTAA AACCCCTCAC AATATTACAT 3120
 ATGAAGTGA GACAGTCTAT ACAAGTACTG ACTATGCTTT GTTTAGGTGG ATGAAATCTC 3180
 TCGGCGTTTA ATTTGGGAAA AAAACCTGAA GTATATTCC ATCCATAACC TTGAGGCTTC 3240
 TCTTGGTGC CATACATATG AACTGGCTAT GAACCACTG GGGGACATGG CAAGTATAGC 3300
 TTCAGTCTCT GTCCCACTG CACCATTTGC TTTAGTTCC TGCTGATGCC TGGCCTCTTT 3360
 CTTCCTTGTG TTAGACCAGT GAAGAGGTGG TTCAGAAGAT GACTGGACTC AAAGTACCCC 3420
 TGTCTCATTG CCGCAGTAAT GACACCTTT ATATCCAGA ATGGGAAGGT AGAGCCCCAG 3480
 ACTCTGTGCA CTATCGAAG AAAGGATATG TTAATCCTGT CAAAATCAG GACTCTCTCT 3540
 TTCTTCTGGG TGTGCATATG TAATCTGGCA TGACCTTTTC CTTTTTCTGC TGCTTTGTTC 3600
 TTGAGGTGAA AGGGCACCAG GAAAAGAGGG CAAGGAATTA AGGTACATCT CCCCATTCCC 3660
 ATTCTGTTAT TTAACCTCAT TTGTTCTGT ACATTTGGGT TGTTCCTGGT TTTTCTTTTT 3720
 CTTTTCCCTT TTTTTTTTTT TTTTTTTTTT GAGATAGAGT CTCACTCTGT CGCCAGGAT 3780
 GGAGTGCAGT GGTGCAATCT TGGCTCACTG CAACCTACAC CTCCCGGGT CAAGCGATTG 3840
 TCCTGCCTCA GCCTCCTGAG TAGCTGAGAT TACAGGCACG CGCCACTACG CCTGGCTAAT 3900
 TTTTCTATTT TTATAGAGAT GCGTTTTAC CATGTTGGCC AGGCTGGTCT TGAAGTACC 3960
 TCAGGTGATC CACCTGCCCT AGCCTCCCA AGTGTGGGA TTAGAGTCAT GAGCCATGCG 4020
 GGCCTGGTTT TTCTTTATTA CAAATAGTGT TGCAATAAGC ACCCTGTGTC ATATGTTTTT 4080
 GTGCACATGT ACAAATATTT ATGCAAAATA AGTCCTAAAA TTGGAATTGT TAGGTCACAA 4140
 ATAATCCTTT CCCCCCCCC AAATTTTTTT TTTTTTTTTG AGACAGCGTC TCTGTCACCC 4200
 AGGCTGGAGT CCAGTGGCGC AATCATGGCT CACTGCAGCC TCAACGTCTC AGGCTCAAGT 4260
 GATTCTCCAA CCTCAGCCTC CCTAGTAGCT GGGAAATAGA AGCACATGCC ACCACACCCA 4320
 GCTAATTTTA AAAAATTTTT TGTTAGAGAC AGGGTTTTGC CATGCTACCC AAGCTGGTCT 4380
 CAAATCTCTG GGCTCAAGCA ATCTGCCCGC TTGGCCTCC CAAAGTGCTA GGATTACAGA 4440
 CATGAGCCAC CATGCCAGC CCAAAAAAGT TTTTGCAATC TTACATTCTT ACTAGCATGA 4500
 GAATGTCAGT TTTTTCACAA CCCAAACAAC ACAGGATTGT ATCAGCAAGA TAAACAATTG 4560
 ATTTAAGTGT CATTTAACA ACACCTTTTG ACCCCAGAA CCTACCAGAT GCAGTGTAG 4620
 GCAGCAGAGA CTCAAGATGA CTAAGACACA ACCTGTGTCC TCAGGAAATC TCAATCTAAA 4680
 AAAATAGAAC AGGAAAGAAA GAAAAATCTA CAATCTAGCT GCACAAACAA TAATAGCTAA 4740
 TACTTTTTGA GATTTTATTG TTTGTCAGGA ACTTCTTAAC TCTTTACATG AGTTTAAATA 4800
 TTTAATCCTT TATAACAATA TTTTATGCAT AGAGAACTG AGACACAGGC AAATTTAGTA 4860
 ACTTACCCGG GGTACATAG CTAATGGGTG GCAAAGTCAG GGTAGCTCC CAGGACAAAT 4920
 GCCTCCACAG CTGGTACTGT GCTCTGCTTT ACTGTAGCTA ATAGTAAAAA TGGTAGCAAA 4980
 AATCAATAGC AGTAGAACAG TGCAACAGAT ATTAAGCGGA AGAGGAAGAC TCACAACAAT 5040
 GACAACATTT GTGCTGAAAT TTTTAAAGAC ACATGGAATT TCCTTCAGCC GGTAGAGAG 5100
 AAGATATAGA AATGTAAACA CCAAGATTG ATAGTTTCTC TGTATCCCTT TCAGGGTCAG 5160
 TGTGGTTCCT GTTGGGCTTT TAGCTCTGTG GGTGCCCTGG AGGGCCAAC CAAGAAGAAA 5220
 ACTGGCAAAC TCTTAAATCT GAGTCCCGC AACCTAGTGG ATTGTGTGTC TGAGAATGAT 5280
 GGCTGTGGAG GGGGCTACAT GACCAATGCC TTCCAATATG TGCAGAAGAA CCGGGGTATT 5340
 GACTCTGAAG ATGCCTACCC ATATGTGGGA CAGGTGAGAT TGCTCCACAC AATTATACAG 5400
 CTCTGTGGC TCCTCCTTCC CCAGCATGAT GTTTGTACT GGAACAATT CCAGAAATAC 5460
 TGTCTTCTGT TATCCTATCC TGCTTCTTG ATGGAATAAT TTCCACAGA AGGCCAAGAA 5520
 GATTTCCACA ATCTGGGGGA ATTTAGGGAG CTTAAGCTAC TATAGCTCCT ATTTGCATCT 5580
 CTGCCATGGA GAGAAAACAG AGGCTAGGCT ACCTACCCCA TAGACTTCCG AGCTGGGTTC 5640
 TATAACCTC TGCTCAATT CCACTCCCA CAACAAACCC ACAAAACCC CATGCTATTT 5700
 TCACAAATG TGTGGCTTTA TTTTATATGA TCTCAGTGTG AGTTTTCAGA ACATTTCAGC 5760
 AAATTATGTA AGTTTACATG CTAACATCTA TAAATGAGA GAAAAACAA GTTGCTTCAT 5820
 ATAAGAGATA AGGATTAAC TCAGTTCCTC CTGCATGATC CTCTAGTCAT AGGAAGGAAA 5880
 TCATATCTGA AAGGGAGGCA ACCTGAGGGG TTTTTTATAC ACATAGGGCT GGGTCTGATA 5940

GACAAATATAA TGTAGGGCCT TCACAACAGA AACCTCTGAA ACAGGGACAG CAAGTTTGAG 6000
AATAAAAAATG ATGGCTACTG TGTCTAAGC CGTGTCCTTA GTGCATTTTT TCTTTTCTT 6060
TTTTTCATTT AATCTCATAA CAACTCTGTT AGGTAGACTT ATCTTGAATG TATAGGTGAG 6120
GAAATGGACA CTAAAGGAGA TAAGACAGTA TAATTCATAC CACTAGTATG TAACAATGTA 6180
AGATGTATCT ACCAGGGATG TTTATCTTCT GCAAACATTC CTAGGTATAT CTCCCATGCA 6240
CATGTGCAAG AATTTCTTAC TAGGATATAA TGCCTTGGAA CTGAATTGTC TGGGTCTTAG 6300
GGTATGTCTG TCTTCAACTT TACTACACAA TGTCAAATTG TTTGCCAAAA TATTTGGAAA 6360
AATTTATACC TGCAATGTGT AAGAAATCCC CTTCAATCAC CTTTTTATCA GTATGTTTAT 6420
CTGGCCATTT GCATTCTTTC TTCAGTGAAT TAACTGTTTT TATCTCTTGC TCATTTGTTT 6480
TTCTTTTTAT TTTTTTGAAG TAGGGTCTTA CTCTGTTGCC CAAGGCTGGA GTGTGGTGAT 6540
ACAGTCATAG CTCACTGCAG CCTCCACTTC CGGGCTCAAG CAATCCTCTC GCCTCAGCCT 6600
CCCAAAATAGC TAGGATATAG GTGCATGCCA TCATGCCAC CAATTTCAAA AAACCTTTGA 6660
AATTTTTTTT TTGTAAGC TAGGCATGGT GGCTCATGCC TGTAAATCCA GCACTTTGGG 6720
AGGCTGAGGT GGGGURCNTN UDAGGATCGC TTGAGCCCAG GAATTGGAGG TCGGCCTGAT 6780
ACAACATAGC AAGACCTCAT CTCTACAGAA AAAATTTTAA AAAGTAGCCA GGTATGATGG 6840
CGTGCAATAG TCTAGCTACT CCGGAAGCTG GTTGGGAGGA CAACTTGAGC CTGGGAGTTC 6900
AAGGCTGCTG TGAAGTGTGA TCATGTCACT GCTCTCTAGC CTGGGTGACA GAGTGAGACC 6960
CTGTCCCAA AAACAACAAC CGTTTTTTTT GGTAGAGACA TTGTCTCGCT ATGTTGCCAA 7020
GGCTAGTCTC AAACCTCTGG GCTCAAGCAA TCCTCCACC TCCCAAAGTG CTGGGATTAT 7080
AGATGTAAGC CACCATGCCT GGCTACCCCT TTTTTTTTTT TTTTGAAATG GAAGTTTTCG 7140
TTTTGTCAAC TAGGCTTGAG TGCAGTGGCG CGATCTTGGC TCACTGCAAC CTCCACCTCC 7200
TGGATTCAAG CAATTCCTCT GCCTCAGCCT CCTGAATAGC TGGGATTATA GGCACCCGCA 7260
ACCACGCCCG GCTAGTTTTT GTATTTTATG TACAGACAGG GTTTCACCAT GTTGGCCAGC 7320
TGGTCTTGAA CCCCTGACCT CAGGTGGTCC GCCCGCCTCG GCCTCCCAA ATGCTGGGAT 7380
TAAAGTGTG AGCCACCATG CCCCAACCCCT TACTCATTTT TAATTGGATT GTTTTTTCTC 7440
TTTCTTAGCG ATTCTTAAAA GTTTAAAGAG AATATTTGGA TACAATACTA TGTATTTAAA 7500
AGTTGAGGTC TGTCTTTCCA TTCTTCTCAC GATGTCTTTC AATCTAGAAA AGTTAATTTT 7560
AATAGGCTG GCGCGGTGGC TCACGCTTGT AATCCCAGCA CTTTGGGAGG CTGAGATGGG 7620
TGGATCAGAA GGTCAAGAGA TGAAGACCAT CCTGGCTAAC ATGGGTGAAA CCTGTCTTCT 7680
ACTAAAAATA CAAAAAATT AGCTGGGCGT GGTGGCAGGT GCCTGTAGTT CCAGCTACTC 7740
GGGAGGCTGA GGCAGGAGAA TGGCGTGAAC CCGGGAGGTG GAGCTTGCAG TGAGCCGAGA 7800
TTGCACCACT GCACTCCAGC CTGGGCAACT GAGCAAGACT GCGTTTCAAA AAAAAAAAAA 7860
GTTAATTTTA ATATAGTAAA ATTAGTAAAA GGATTAATTT TCCCTTTGCA ATTTTGTGAA 7920
TGTGTTTTAT TCGTTTATGA ATGGAGAAAG GTAAGAAAAA ATAAATTTA AAAAAAGA 7980
GATGTGGCCA GGTACGGTGG CTCACACCTA TAATCCAGT AGTTTGGGAG GCTGAGGCAG 8040
GCAGATCACT TGAGGTCAGG AGTTTGAGAC CAGCTGGGAT AACATGGTGA AATCCATCT 8100
CTACTAAAAA TACAAAAATT AGCCAGGTGT GATTGCGCAC GCTTGTAATC CCAGCAGGCT 8160
GAGGCAGGAG AATTGCTCGA ACTCAGGAGG CAGAGGTTGC AGTGAGCCAA GATCATGCCA 8220
TTGCACTCCA GCCTGGGTAA CAGAGACTCT GTTTCAAAAA AATAAAAAA TAAAAAGGA 8280
AGAGATCTGA TAGGGCGGCC AGATAAACAT TTTAAAGGGG ATGGTATTAT AAGTTTGTTC 8340
CCAGCATAAT GCCAGGTAT TCTGACTTTA AAGTATCATC ACATAATATC TTTTGTAGTC 8400
AATTTCCAAG ATATTCTGTT TCACTTGTA TCTGTGTAA TTTTGGCAC CAGGAGGCAT 8460
CAGGGATTG GAGCACATGG CAGAAACAAA GGCATCTTGA AAAATATCAA GGCAGTAGAC 8520
CACTGTAATC TTAATATGGC ATATCAAATG CTGCTATTGC TGTTAATATT TAGATAATGT 8580
TAGATAATGT ATTTTATAG AGGGTATCTC ACTATCTTGC ACAGGCTGGA GTAGAGTGGC 8640
TATTCACAGC ATGATCACAG TACATAAAG GCTCAAATC CTGGGCACAA ACAATCCTCC 8700
TGCCTCAGCC TGCTGAGTAG TAGATAATAA GTTCTTGTGG ATGCAACCTT AGGGTTCTGA 8760
AGGGGTAGTC TGTAGGAAAA TGAATTGCTG AAAAGAATAC ACCACCTTAA CATGGGCTAT 8820
TATTCGATTC CATAATTGTG GCTTGCCAAT GAAACATTGC TAACTACCTG TAAAAATAG 8880
TGTTGGAAGT CATAGGCTAA ATTGCTAAGT TCTTTAATCT ATTTTAGTGT CTGTTATGT 8940

ACTTTTATAT TTGTCTTTG ATGAGAGCAC AAGGATCACA CCAGTTCCCC TGATATAGGT 9000
 GCAGAGGGCC CAGGTCTTCC CTCTAGCTAA GCCTTGGCCT TGGCCTCCTA CCCACACAGC 9060
 AGCTGGTGCC TTCCTGCCCC CTGAGGCTAA TACATACTAT GTGGCCAGAA GATGGTTTAT 9120
 GCTTTTAAAA AAAATCTTAT TTCAGAAATC TTTCCTACT GTTTTCTCC CACATTTATG 9180
 TCTTAAAAA CCTGTAGGGG ATTTTTTTTT TTTTTTTTTT TTTGAGATGG AGTCTCGCTC 9240
 TCGCCAGGC TGGAGTGCAA TGGGCGGATC TTGGCTCACT GCAAGGTCTG CCTCCAGGT 9300
 TCACGCCATT CTCCTGCCTC AGCCTCCCCA GTAGCTGGGA CTACAGGCGC CCGCTACCAC 9360
 GCCTGGCTAA TTTTTTGCA TTTTAGTAG AGACAGGGTT TCACTGTTAG CCAGGATGGT 9420
 CTCGATCTCC TGACCTCGTG ATCCACCCTC CTCAGCCTCC AAAGTGCTGG GATTAACAGG 9480
 CATGGAGCCC CACCGCACTG GCCTGTATTT GTGAGGAAGA ACAGACCCTC TTTAGAAGCC 9540
 CTAGACTGCT GCCTCTGTTA GTTCACTGGC ATCACTCAAA ATATTGGTTG AGTTTCTTAC 9600
 TCACTGAGTT GGTTTTTATG TGTGGTGAA GCGGGGAATC CTCTTTTCAT ATTCGTTCTC 9660
 ATTGCCTATT GCTTTGTCTC AGTCCTATTA CAATCTTGTT TCTTCCAGGA AGAGAGTTGT 9720
 ATGTACAACC CAACAGGCAA GGCAGCTAAA TGCAGAGGGT ACAGAGAGAT CCCCAGGGGG 9780
 AATGAGAAAG CCTGAAGAG GGCAGTGGCC CGAGTGGGAC CTGTCTCTGT GGCCATTGAT 9840
 GCAAGCCTGA CCTCCTTCCA GTTTTACAGC AAAGGTAAGA AGCTGCTGAT CCTATACAGC 9900
 ACTGTCTTTT ATGATACAAA CTGATGGTT TCTGAAGGA CCTTGGGTAT TTTCAGTACT 9960
 TAGTTTTTGT ATTCACATGG AGGTGGCCAG AGAGAAATTA ACAACTGCTG CAGTATGGAG 10020
 CAGCATCTCT GTGGTAAACC CTCTGACAC GGATGGAATT CTTCAAACAG TCTCCTAGAC 10080
 TGGGAGATCC CACAGGGTGA CCCTTGGATT GCATAGAGCC TCACGCTGGT AGTTTGTATT 10140
 CTAGGTGTGT ATTATGATGA AAGCTGCAAT AGCGATAATC TGAACCATGC GGTTTTGGCA 10200
 GTGGGATATG GAATCCAGAA GGGAAACAAG CACTGGATAA TTAACACAG GTAATGATGG 10260
 GAACACTACT TTTGTTATTC AGTCACCCTT TTAACACTCA ACCTCACCTC CAGCTTCCCG 10320
 ATATTCTTTT CTCTGTCCA AATCAAGAAA AAATTATCTC AGAGTTCTCA CTTCTATCTT 10380
 CTCAGTCAGA GGCTCTTAAT TCTCAGTCTG AACTTAATG GOCAGTGTGT TAGTCCATTT 10440
 TGCATTGCCA CAAAAGAATA CCGAGACTG GGTAGTTTAT AAAGAAACGA GGTTTGTTTG 10500
 GGTATACAAA GCGTGGCACT AGTATCTGCT CAGCCTCTGA TGAGGCCTCA GAGCTTTTAC 10560
 TCATGGCAGA AGGCAAAAGA GGGAGCAGGC ATGTCACATA GTGAGAGAGG GAGCAAGAGA 10620
 GAGAGGGAGG TGCCGACTCT TTAAGAACC AGCTCTTGCA TGAACATAA GAGTGAGAAC 10680
 TCACTCATCA CCAAGGGAT GGCACCAAGC CATTCCATGA GGAATCCACT CTCATAACCC 10740
 AAACACCTCC CACTATGCC CACCTCCAC ATTGGGGATC ACATTTACAGC ATGAGACTGG 10800
 GAGGGGACAC ACATCCAAAC CATATCCGCC AGACAATAGT GCTCAATTAT GTGCTGGGCA 10860
 GATGCTCCCT GTGTGCAAGG TGCTTAGTGA CATACATAAA CCAACGAGCA GATGACACCT 10920
 TCAGTGAGCT CAGAGCCCAA TAAGACAGAC CTAACCTAACC ATGAGATAAA GCAGTACAAA 10980
 GAACCAGCAG GAGCTTGGGA ATTACGTATT TTTACTTTCT TTTGTCTCTA ATGTGATCAG 11040
 TTTCTTAGAT GGTTCATT AGCAATCTGT CTTTAACAGT AGGGGAGCAG CGTTAAAGGT 11100
 TTAATATTCC TTTTGAACAG TTTTTCCT TCAAAATACA CTTAAGATAC ACGTATATAA 11160
 GAACTTGCCA AAGATTGTGA AGAGAAACAT TTTTAGAAA TAAGATATAA AAAAAAAG 11220
 TTAGTGTTAC TTTCTATGT TGGGAACAA AGAAACTCC AGGGTACCTT GCTTCCATT 11280
 TCTCTTAGC ACCTTGAGC TTTTGGGAG GGGCAGATTG ATAACAATTA TAGTTTCTC 11340
 TTCTGGCTG ATCACCATTA ACCTGGCAGC AGCACTGGCT AAATCTCCTG TCCTTAGTGC 11400
 CCTCCAAGGA GCAGGAGCCC TAGACTCTGG GTCGCTGACA GACTCACGCA GTGGTGTTGT 11460
 TCAAACTGA AGCAACTTTT TATATCACAG TTCCAACCTA AGGTGAACCT GAGCATCTTC 11520
 CCAAGTCTCC CACAGCTTCT GTCTGTGTT GTCCCTTCTC TTGACTCCCA GGTCCAAGCA 11580
 CTTACCTGT TCTTTCATGA TCAGGTACCA TGTGTGAGA TAGCTTCCAA GAGAGCTGGG 11640
 AGGAAGAAAG GACACACCCG GGCAGGATCA GGAACACTGG GGGCCCCTGG AGAAGGGGAG 11700
 AGTGGGGGAG GGTACAGGTT TTAATAAAA TGTGTTGGTA ATTAGAGAAT TGCTGGTTGG 11760
 GGAAAGAGGT CTGAAACAA TTCAGGAAGA TAAACAAGAC AATCTCTCCT CTCTCCTCTT 11820
 TCTCAGTCG TCTCTTGT CTCTAGTCT CGCTACTCAT TTCCTTAGTA ATCTCATCCA 11880
 CTCTCATAGT TTCATCCATC TCTCCTATGG GGTTAACCC CAAATCAAGA TCACCAGCTT 11940

CAGCCTCCTT CTTATGCTCT AAACCTCACAT TTTCAAGATT AATATTCCCC AAATACAGCT 12000
 CTGATCATAT CACTCTCCCA CTCAAAATCC CTCACGGCT CCTCACGATG ATGGGTCACA 12060
 GAGTAAAGGT GAAGCTTTTT AACCTTGCG TAAAGGTAAT TCAACCTGAT CTCAATCTGC 12120
 CTTTCAGAC ATCTCTCCCA CTACACCTG TTAGGCACAC TGCTTTTCAG CTACATGATC 12180
 CTAACAGTGC CCCACACTTT CCTGCCTCTG TTGTTTCATTT CACACCCTTC CACTGGCATC 12240
 CCCTTCCAC AGGTGCGAAT TCTACTTAGC CTTTTGGCTC AGCTCAAATG CCACCTCTTA 12300
 CATCAAGCCT CTAAGATTCT CTTGATCAGA AGGAATCTTT CCCTCCTTG ATACCTACAG 12360
 TATTATGCCT TCTCCTATT TCTTGACTTT AAACCTTTA AAGTTAAAA ACATCATATT 12420
 CATTTTTGTG TACCATCAGT ACCTCGCACA ATACTCAGTA AATATTTTAA TGAATAAATA 12480
 AACTGAGAGT ACTAAGTATT TTTCTTGATT GGTCTTACAG CTGGGGAGAA AACTGGGGAA 12540
 ACAAAGGATA TATCCTCATG GCTCGAAATA AGAACACGC CTGTGGCATT GCCAACCTGG 12600
 CCAGCTTCCC CAAGATGTGA CTCACGCCAG CCCAAATCCA TCCTGCTCTT CCATTTCTTT 12660
 CCACGATGGT GCAGTGTAAC GATGCACTTT GGAAGGGAGT TGGTGTGCTA TTTTGAAGC 12720
 AGATGTGGTG ATACTGAGAT TGTCTGTTCA GTTCCCAT GTTGTGTGC TTCAAATGAT 12780
 CCTTCTACT TTGCTTCTCT CCACCATGA CCTTTTCCA CTGTGGCCAT CAGGACTTTC 12840
 CCTGACAGC TGTGACTCT TAGGCTAAGA GATGTGACTA CAGCTGCCC CTGACTGTGT 12900
 TGTCCAGGG CTGATGCTGT ACAGGTACAG GCTGGAGATT TTCACATAGG TFAGATTCTC 12960
 ATTACAGGGA CTAGTTAGCT TTAAGCACCC TAGAGGACTA GGGTAATCTG ACTTCTCACT 13020
 TCCTAAGTTC CCTTCTATAT CCTCAAGGTA GAAATGTCTA TGTTTTCTAC TCCAATTCAT 13080
 AAATCTATTC ATAAGTCTTT GGTACAAGTT TACATGATAA AAAGAAATGT GATTGTCTTT 13140
 CCCTTCTTTG CACTTTTGAA ATAAAGTATT TATCTCCTGT CTACAGTTTA ATAAATAGCA 13200
 TCTAGTACAC ATTCAATTTG TGTGGATAC TGTGTTAGGT GCTGGAGGAA AAAAGATGAA 13260
 TAGAACATCT TCTATGACT TGATGCGCTC ACAGTCTGGT TGTAGAGACT GTCACATAAA 13320
 CATTTTATCC CAATTCATTT ATTTGTTTAT TCCTTCAGCC AATATATATT GAGTTCTTAC 13380
 TCTGTGCAA GAACTGACT ACATTCTGG GATTAAGTGG ATATAAGGAG ATCTCAGTGT 13440
 TTAATCTGCC TGAGGGGAGA CTAATTAAG TGACATGGAA ACTTGGGTCT TGA AAAACAT 13500
 TTTAAGTTA TTTTCTCTT TCTCTCTC TCGCTCTGTC TTTCTCTC TTTCTGTCAG 13560
 GTCTCCTCT GTTGCCAGG CTGGAGTCAG TGGCACTCAT AGCTCACTGC AGCCTTGATC 13620
 TCCTGGGCTC AAGAGTTCTT CCCACCTCAG TCTCCTAAGT AGCTTGAGT ACGG 13674

【0160】(2) 配列番号：2の情報：

(i) 配列の特徴：

(A) 配列の長さ：1108塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：2：

GCTTTGGCTC CCAAAGGCCT GGGATTACAG GCGTGAACCA CTGCGCCTAG CCTGTTAGCA 60
 GCTCTTAAAA TCCAGAGGCA TAAGCCTGTA TTTTGAGGG TTTATGCATG GAATCCAGCT 120
 AGAAACTGAG TCTATTACAG ATCCCATTTA TTATCCTTTC TATTCCAAGA AGCCTTTTTT 180
 TCTCCTTCCC CACATCTGTT TATGGAAGAA AATGAAGTTT GGGGTGTGGT TTGAGGAATC 240
 AGCTAGATTC TTATGATCTG TCACATGCTT GGATGTTGGG GAAGCATTTG GAGAAGCTCA 300
 TGTGACTTGT CCTAGATTGG GGATTTTAAT TGAGACAGAT GATGTTTATC GGGCATCCCA 360
 CCACCTGAGA GTTTTAGCAA CAGAGTCACA TGTGAGTCCA TCAGAACTTA CGGCATTGAT 420
 TCAAGTGCTG TCATAAATAA CCAGGACTGC TGTTTTGGT TACTTTTAAA GACAGTTTCA 480
 TCTGGACTTT CTGGGCATAT CCTCCTTCAG CAAAACCACA TTAGGCTGGG AAAACTATTC 540
 TGCCCTGGAAG TAATGACAAC TTGCAACCAA CAAGCTTATA AAAATACAAA GAATCTGGA 600
 GCCTATGGCT TCCATTACAT TATTCTTTTA TAGCCTTTTA TGTTCATTAC CGCATCCCAG 660
 AGGTGAGAGT CAGACACAAA TATGAAAATA GGTTCATG TTGGAGAGGT AAATCCTAAC 720
 AGGAAAGGGG TAGGAAAAGA TATAATCCCC CAATATTTAA ATAAAGATAT TGAAGAAGAA 780
 GGATGGGAGA GACTAGGGCT GTGTCTTCC TTTTACTCAC CAAAAGAGAA AGTAAGCTCC 840
 TATTTGAGTC AATAGATATT GAGGTCTGT TATTTGCCAC CAAAGACAGT CTGTGAGAC 900

TAAATAGCTA GTAATTCCT ACCCTGGCAC ACATGCTGCA TACACACAGA AACACTGCAA 960
 ATCCACTGCC TCCTTCCTC CTCCTACCC TTCCTTCTCT CAGCATTTCT ATCCCCGCT 1020
 CCTCCTCTTA CCCAAATTTT CCAGCCGATC ACTGGAGCTG ACTTCCGCAA TCCCGATGGA 1080
 ATAAATCTAG CACCCCTGAT GGTGTGCC 1108

- 【0161】(2) 配列番号: 3 の情報: (ii) 分子の型: ゲノムDNA
 (i) 配列の特徴: (iii) 仮定: なし
 (A) 配列の長さ: 48 塩基対 (iv) アンチセンス: なし
 (B) 配列の型: 核酸 (v) 断片の型:
 (C) 鎖の数: 一本鎖 (vi) 起源:
 (D) トポロジー: 直鎖状 (xi) 配列の記載: 配列番号: 3:

CACACTTTGC TGCCGAAACG AAGCCAGACA ACAGATTTC ATCAGCAG 48

- 【0162】(2) 配列番号: 4 の情報: (ii) 分子の型: ゲノムDNA
 (i) 配列の特徴: (iii) 仮定: なし
 (A) 配列の長さ: 1427 塩基対 (iv) アンチセンス: なし
 (B) 配列の型: 核酸 (v) 断片の型:
 (C) 鎖の数: 一本鎖 (vi) 起源:
 (D) トポロジー: 直鎖状 (xi) 配列の記載: 配列番号: 4:

GTAACGTTTG CAACTTCCTA GATCTTTTAG CTTTTCATTC CTGTCAATTC TCTGAGTATT 60
 AGGGATGTAG TGACTTGAGG ATCACAATAA ACTTTTAGCC TCTGCAGATG AAAACAGAGA 120
 TGCATCTCTT AGGTCATTCC CTGGCTAAAT AAAATCTGCC TGGAAATCTG TAGAATTCCT 180
 TGTATGATTT ATATATATAC ATACATGATT GTTAGTAAAA GCAAAGTATA TAGGGAATCA 240
 TTTCCCATC CTCAAGAGT GGCCTTTCTG CAGTGTITTC TACTTTGGCC AACAAGGATC 300
 AAAACGGTTA ACTCCTTAGT GAGGAGGAGG AGAGTGGTAT GGGGAGGTAG TAGCTCAGTG 360
 CTTCTGTTC ACTGAGACAT CTCAAAGCCC TTAACACTCT AGTTTTTAAA TGTCTACTG 420
 GACATTTTGC CAGTTTGCAA AATTACATGT AAATGGACTA TAAGCAATTG TGTAAGCCAT 480
 ATGTCATGCT GCAGGCTGCA AATTGTCTT AAAATGGAGG ATTTGTAATT AAGAAAGCCA 540
 ATGCAAGAAA TGAGTGAAGC TAAGTAGAGT AAACCTATGA AAAGCTGTGA ATTTATCAT 600
 CATAGAACAT TGCTTTTCAG TCTGAACATT CTTCTAACAA ACCTTGATC TGAGGCTTCT 660
 TGTCTTTGC GGCAGCCACA GTGGGTTTTT GTTGTAGGG GAAAAATAAA AACCTTGCCC 720
 GCAGCATCTG GTTAAGATTA GGGCAGTTTC CTGCCTAAGG AGGGAAGGGA GAGAAAAAGG 780
 AAGAAGAAAT GCATAAGGAG AATGAGGAGA TATACAATGT CTCAGAAAAC AGGAAACATT 840
 GTCCTATTTT CCCTTGTCTT CTTCTGACAA GATCTGGGAA AGTACCAGAA TTTAGGCACG 900
 AAAGAGAAGA ACGCCTCGAA GAAATGATCA GGAAGCAAAA CTTAGACGGA AATCTCTCCT 960
 TTGTGTATTC TGAACCCAC TACCACCTTG CTATTTGTCT GTCTCCAAGC CTGCTAGGGA 1020
 CCTGGAGGA AAGCACTGA GCCATTCTG ATTGTCCAGT TTCTATCCC CATTTCTGGT 1080
 TGTGTACGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGAGAGAGAG 1140
 AGAGACAGAG AGAGAAACAG AGAGAGTGTG TGTGTCTAA ATCTCCCGAG AGAGAGAGAG 1200
 AGAGAGAGAG AGAGAGAGAG AGAGAGAAA GAGAGAAATG GCTAAATCCC CCTAGATCAA 1260
 AGTCCTTGGA ACCAGATGTA CCAGCATCCT ATCTAAACAC AGGCCCTCC TGAATATCAT 1320
 TGTTTTATCA CCTTTTTC GTCTACCTTT CTCTCCTCA TAAAGCCTAG TTTTCTCTG 1380
 TTTCCCTGCC AAATGGAAGA GTTTCCCTA ACTACATTCT TCTGCAG 1427

- 【0163】(2) 配列番号: 5 の情報: (ii) 分子の型: ゲノムDNA
 (i) 配列の特徴: (iii) 仮定: なし
 (A) 配列の長さ: 121 塩基対 (iv) アンチセンス: なし
 (B) 配列の型: 核酸 (v) 断片の型:
 (C) 鎖の数: 一本鎖 (vi) 起源:
 (D) トポロジー: 直鎖状 (xi) 配列の記載: 配列番号: 5:

GATGTGGGG CTCAAGGTTT TGCTGCTACC TGTGGTGAGC TTTGCTCTGT ACCCTGAGGA 60
 GATACTGGAC ACCCACTGGG AGCTATGGAA GAAGACCCAC AGGAAGCAAT ATAACAACAA 120
 G 121

【0164】(2) 配列番号：6の情報：

(i) 配列の特徴：

(A) 配列の長さ：462塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：6：

```
GTGCCTGGGG TCCTGGAGGG GGCATGGCAG GAAGGCTGAG ACCTGAGCTC TCTCATCTTA 60
GCTTCCAGAC TCCCTTCTTC AATCCAAATG CTTTATTCCA AGCAAATCAG TCCCTCTTCC 120
CTAACTCATG TTAACATACG GTTTTCATTG CTATGCTTCA ATCATCCTCT TGTCAAACCT 180
GTATTCTTC CTTTGGGTTT TATAAGTGTG TAACATTCTT CTTTGGGAA GAGTCCCAAG 240
ATTAATGCTG TTAATCCATA AGCAATTTTT CTGTCTCTCC AGAGCTTGTG TGGTTGTTTA 300
CATATTATCT CTCCTCTTGC AGGCTCTTAA TTCCATGGTT AGTTCCCAAA CTAAACTGTA 360
AACCTTTATG ATTGTGAGTT TCCTTTATTC TCCTAAAACC CTTCAACAATA TTACATATGA 420
ACTGTAGACA GTCTATACAA GTACTGACTA TGCTTTGTTT AG 462
```

【0165】(2) 配列番号：7の情報：

(i) 配列の特徴：

(A) 配列の長さ：123塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：7：

```
GTGGATGAAA TCTCTCGGCG TTTAATTTGG GAAAAAACC TGAAGTATAT TTCCATCCAT 60
AACCTTGAGG CTTCTCTTGG TGTCCATACA TATGAACTGG CTATGAACCA CCTGGGGGAC 120
ATG 123
```

【0166】(2) 配列番号：8の情報：

(i) 配列の特徴：

(A) 配列の長さ：85塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：8：

```
GCAAGTATAG CTTGAGCTCC TGTCCACCT GCACCATTG CTTTAGTCC CTGCTGATGC 60
CTGGCCTCTT TCTTCTTTGT CTTAG 85
```

【0167】(2) 配列番号：9の情報：

(i) 配列の特徴：

(A) 配列の長さ：156塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：9：

```
ACCAGTGAAG AGGTGGTTCA GAAGATGACT GGAATCAAAG TACCCCTGTC TCATTCCCGC 60
AGTAATGACA CCCTTTATAT CCCAGAATGG GAAGGTAGAG CCCAGACTC TGTGACTAT 120
CGAAAGAAAG GATATGTTAC TCCTGTCAA AATCAG 156
```

【0168】(2) 配列番号：10の情報：

(i) 配列の特徴：

(A) 配列の長さ：1624塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：10：

```
GTACTCTCCT TTCTTCTGGG TGTGCATATG TAATCTGGCA TGACCTTTTC CTTTTCTGC 60
TGCTTTGTTT TTGAGTGAA AGGGCACCAG GAAAGAGGG CAAGGAATTA AGGTACATCT 120
CCCCATCCC ATTCTGTTAT TTAACCTCAT TTGTTTCTGT ACATTGGGT TGTTCCTGGT 180
TTTTCTTTTT CTTTCCCTT TTTTTTTTTT TTTTTTTTTT GAGATAGAGT CTCACCTCTGT 240
```


CGCCCAGGAT GGAGTGCAGT GGTGCAATCT TGGCTCACTG CAACCTACAC CTCCCGGGTT 300
 CAAGCGATTG TCCTGCCTCA GCCTCCTGAG TAGCTGAGAT TACAGGCACG CGCCACTACG 360
 CCTGGCTAAT TTTTCTATTT TTATAGAGAT GCGTTTTCAC CATGTTGGCC AGGCTGGTCT 420
 TGAAGTACC TCAGGTGATC CACCTGCCCTC AGCCTCCCAA AGTGCTGGGA TTAGAGTCAT 480
 GAGCCATCGC GGCTGGTTT TTCCTTATTA CAAATAGTGT TGCAATAAGC ACCCTGTGTC 540
 ATATGTTTT GTGCACATGT ACAAATATTT ATGCAAAATA AGTCCTAAAA TTGGAATTGT 600
 TAGGTCACAA ATAATCCTTT CCCCCCCCC AAATTTTTTT TTTTTTTTG AGACAGCGTC 660
 TCTGTACCC AGGCTGGAGT CCAGTGGCGC AATCATGGCT CACTGCAGCC TCAACGTCTC 720
 AGGCTCAAGT GATTCTCCAA CCTCAGCCTC CCTAGTAGCT GGGAAATAGA AGCACATGCC 780
 ACCACACCCA GCTAATTTTA AAAAAATTTT TGTAGAGAC AGGGTTTTGC CATGCTACCC 840
 AAGCTGGTCT CAAATTCCTG GGCTCAAGCA ATCTGCCCGC TTCGGCCTCC CAAAGTGCTA 900
 GGATTACAGA CATGAGCCAC CATGCCAGC CCAAAAAAGT TTTTGCAATC TTACATTCTT 960
 ACTAGCATGA GAATGTCAGT TTTTTCACAA CCAAAACAAC ACAGGATTGT ATCAGCAAGA 1020
 TAAACAATTG ATTTAACGTT CATTTAACAA ACACTTTTTC ACCCCAGAA CCTACCAGAT 1080
 GCAGTGTAG GCAGCAGAGA CTCAGATGA CTAAGACACA ACCTGTGTCC TCAGGAAATC 1140
 TCAATCTAAA AAAATAGAAC AGGAAAGAAA GAAAAATCTA CAATCTAGCT GCACAAACAA 1200
 TAATAGCTAA TACTTTTTGA GATTTTATTG TTTGTCAGGA ACTTCTTAAC TCTTTACATG 1260
 AGTTTAAATA TTTAATCCCT TATAACAATA TTTTATGCAT AGAGAAACTG AGACACAGGC 1320
 AAATTTAGTA ACTTACCCGG GGTACATAG CTAAGTGGTG GCAAAGTCAG GGTAGCTCC 1380
 CAGGACAAAT GCCTCCACAG CTGGTACTGT GCTCTGCTTT ACTGTAGCTA ATAGTAAAAA 1440
 TGGTAGCAAA AATCAATAGC AGTAGAACAG TGCAACAGAT ATTAAGCGGA AGAGGAAGAC 1500
 TCACAACAAT GACAACATTT GTGCTGAAAT TTTTAAGAAC ACATGGAATT TCCTTCAGCC 1560
 GGGTAGAGAG AAGATATAGA AATGTAAACA CCAAGATTTC ATAGTTTCTC TGTATCCCTT 1620
 TCAG 1624

【0169】(2) 配列番号: 11の情報:

(i) 配列の特徴:

(A) 配列の長さ: 219塩基対

(B) 配列の型: 核酸

(C) 鎖の数: 一本鎖

(D) トポロジー: 直鎖状

(ii) 分子の型: ゲノムDNA

(iii) 仮定: なし

(iv) アンチセンス: なし

(v) 断片の型:

(vi) 起源:

(xi) 配列の記載: 配列番号: 11:

GGTCAGTGTG GTTCCTGTTG GGCTTTTAGC TCTGTGGGTG CCCTGGAGGG CCAACTCAAG 60
 AAGAAAACTG GCAAACTCTT AAATCTGAGT CCCCAGAAC TAGTGGAATTG TGTGTCTGAG 120
 AATGATGGCT GTGGAGGGG CTACATGACC AATGCCTTCC AATATGTGCA GAAGAACCGG 180
 GGTATTGACT CTGAAGATGC CTACCCATAT GTGGGACAG 219

【0170】(2) 配列番号: 12の情報:

(i) 配列の特徴:

(A) 配列の長さ: 4326塩基対

(B) 配列の型: 核酸

(C) 鎖の数: 一本鎖

(D) トポロジー: 直鎖状

(ii) 分子の型: ゲノムDNA

(iii) 仮定: なし

(iv) アンチセンス: なし

(v) 断片の型:

(vi) 起源:

(xi) 配列の記載: 配列番号: 12:

GTGAGATTGC TCCACACAAT TATACAGCTC TGTGGGCTCC TCCTTCCCA GCATGATGTT 60
 TTGTACTGGA AACAAATCCA GAAATACTGT TTTCTGTTAT CCTATCCTGC TTTCTTGATG 120
 GAATAATTC CCACAGAAGG CCAAGAAGAT TTCCACAATC TGGGGGAATT TAGGGAGCTT 180
 AAGCTACTAT AGCTCCTATT TGCATCTCTG CCATGGAGAG AAAACAGAGG CTAGGCTACC 240
 TACCCCATAG ACTTCCGAGC TGGTTCTAT AACCTCTGC TCAATTCCTC ACTCCACAA 300
 CAAACCCACA AACCCACCAT GCTATTTTCA CAAATTGTGT GGCTTTATTT TATATGATCT 360
 CAGTGTGAGT TTTGAGAACA TTTGAGCAA TTATGTAAGT TTACATGCTA ACATCTATAA 420
 AATGAGAGAA AAAACAAGTT GCTTCATATA AGAGATAAGG GATTAACCTA GTTCCTCTG 480
 CATGATCCTC TAGTCATAGG AAGGAAATCA TATCTGAAAG GGAGGCAACC TGAGGGGTTT 540
 TTTATACACA TAGGGCTGGG TCTGATAGAC AATATAATGT AGGGCCTTCA CAACAGAAAC 600

CTCTGAAACA GGGACAGCAA GTTTGAGAAT AAAATGATG GCTACTGTGT TCTAAGCCGT 660
GTCCCTTAGTG CATTITTTCT TTTTCTTTT TTCAITTAAT CTCATAACAA CTCTGTTAGG 720
TAGACTTATC TTGAATGTAT AGGTGAGGAA ATGGACACTT AAGGAGATAA GACAGTATAA 780
TTCATACCAC TAGTATGTAA CAATGTAAGA TGTATCTACC AGGGATGTTT ATCTTCTGCA 840
AACATTCTTA GGTATATCTC CCATGCACAT GTGCAAGAAT TTCTTACTAG GATATAATGC 900
CTTGGAACCTG AATTGTCTGG GTCTTAGGGT ATGCTGTCTT TCAACTTTAC TACACAATGT 960
CAAATTGTTT GCCAAAATAT TTGAAAAAT TTATACCTGC AATGTGTAAG AAATCCCCTT 1020
CAATCACCTT TTTATCAGTA TGTATTCTG GCCATTGCA TTTCTTCTTC AGTGAATTAA 1080
CTGTTTTTAT CTCTTGCTCA TTTGTTTTT TTTTATTTT TTGAAATAG GGTCTTACTC 1140
TGTGCCCCAA GGCTGGAGTG TGGTGATACA GTCATAGCTC ACTGCAGCCT CCACTTCCGG 1200
GCTCAAGCAA TCCTCTCGCC TCAGCCTCCC AAATAGCTAG GATATAGGTG CATGCCATCA 1260
TGCCCAACAA TTTCAAAAA CCTTTGAAAT TTTTTTTTTG TAAAAGCTAG GCATGGTGGC 1320
TCATGCCTGT AATCCAGCA CTTTGGGAGG CTGAGGTGGG AGGATCGCTT GAGCCAGGA 1380
ATTGGAGGTC GGCCTGATAC AACATAGCAA GACCTCATCT CTACAGAAAA AATTTTTAAA 1440
AGTAGCCAGG TATGATGGCG TGCATAGTTC TAGCTACTCC GGAAGCTGGT TGGGAGGACA 1500
ACTTGAGCCT GGGAGTTCAA GGCTGCTGTG AACTGTGATC ATGTCACTGC TCTCTAGCCT 1560
GGGTGACAGA GTGAGACCCT GTCCCCAAAA ACAACAACCG TTTTTTTTGG TAGAGACATT 1620
GTCTCGCTAT GTTGCCAAGG CTAGTCTCAA ACTCCTGGGC TCAAGCAATC CTCCCACCTC 1680
CCAAAGTGCT GGGATTATAG ATGTAAGCCA CCATGCCTGG CCTACCTTTT TTTTTTTTTT 1740
TTGAAATGGA AGTTTGTCTT TTGTCACTTA GGCTTGAGTG CAGTGGCCCG ATCTTGGCTC 1800
ACTGCAACCT CCACCTCTG GATTCAAGCA ATTCTCCTGC CTCAGCCTCC TGAATAGCTG 1860
GGATTATAGG CACCCGCAAC CACGCCCGGC TAGTTTTGT ATTTTATAGTA CAGACAGGGT 1920
TTCACCATGT TGGCCAGCTG GTCTTGAACC CCTGACCTCA GGTGGTCCGC CCGCCTCGGC 1980
CTCCCAAAAT GCTGGGATTA AAAGTGTGAG CCACCATGCC CCACCCCTTA CTCATTTTTA 2040
ATTGGATTGT TTTTCTCTT TCTTAGCGAT TCTTAAAGT TTAAGAGAA TATTGGATA 2100
CAATACTATG TATTTAAAG TTGAGGTCTG TCTTCCATT CTCTCACGA TGTCTTTCAA 2160
TCTAGAAAAG TTAATTTTAA TAGGCCTGGC GCGGTGGCTC ACGCTGTAA TCCAGCACT 2220
TTGGGAGGCT GAGATGGGTG GATCACAAGG TCAGGAGATG AAGACCATCC TGGCTAACAT 2280
GGGTGAAACC CTGTTCTAC TAAAAATACA AAAAAATTAG CTGGGCGTGG TGGCAGGTGC 2340
CTGTAGTTCC AGCTACTCGG GAGGCTGAGG CAGGAGAATG GCGTGAACCC GGGAGGTGGA 2400
GCTTGCACTG AGCCGAGATT GCACCACTGC ACTCCAGCCT GGGCAACTGA GCAAGACTGC 2460
GTTTCAAAAA AAAAAAAGT TAATTTAAT ATAGTAAAT TAGTAAAAGG ATTAATTTTC 2520
CCTTTGCAAT TTTTGTAAAT TGTTTATTC GTTTATGAAT GGAGAAAGGT AAGAAAAAT 2580
AAAAATTTAA AAAGAAGAGA TGTGGCCAGG TACGGTGGCT CACACCTATA ATCCAGTAG 2640
TTTGGGAGGC TGAGGCAGGC AGATCACTTG AGGTCAGGAG TTTGAGACCA GCTGGGATAA 2700
CATGGTGAAA CCCATCTCT ACTAAAAATA CAAAAATTAG CCAGGTGTGA TTGCGCACGC 2760
TTGTAATCCC AGCAGGCTGA GGCAGGAGAA TTGCTCGAAC TCAGGAGGCA GAGTTGCAG 2820
TGAGCCAAGA TCATGCCATT GCACTCCAGC CTGGGTAACA GAGACTCTGT TTCAAAAAA 2880
TAAAAGATA AAAAAAGGAG AGATCTGATA GGGCGGCCAG ATAAACATTT TAAAGGGGAT 2940
GGTATTATAA GTTTGTTCCC AGCATAATGC CAGGTATTTC TGACTTTAAA GTATCATCAC 3000
ATAATATCTT TTTGAGTCAA TTTCCAAGAT ATTCTGTTT ACTTGTAATT CTGTGTAATT 3060
TTTGGCACCA GGAGGCATCA GGGATTGGA GCACATGGCA GAAACAAAGG CATCTTGAAA 3120
AATATCAAGG CAGTAGACCA CTGTAATCTT AAAATGGCAT ATCAAATGCT GCTATTGCTG 3180
TTAATATTTA GATAATGTTA GATAATGTAT TTTTITAGAG GGTATCTCAC TATCTTGCAC 3240
AGGCTGGAGT AGAGTGGCTA TTCACAGCAT GATCAGAGTA CACTAAAGGC TCAAACCTCT 3300
GGGCACAAAC AATCCTCTG CCTCAGCCTG CTGAGTAGTA GATAATAAGT TCTGTGGGAT 3360
GCAACCTTAG GGTCTGAAG GGTAGTCTG TAGGAAAATG AATTGCTGAA AAGAATACAC 3420
CACCTTAACA TGGGCTATTA TTCGATTCCA TAATTGTGGC TTGCCAATGA AACATTGCTA 3480
ACTACCTGTA AAATATAGTG TTGGAAGTCA TAGGCTAAAT TGCTAAGTTC TTTAATCTAT 3540
TTTAGTGTCT TGTATGTAC TTTTATTTT TGTCTTTGAT GAGAGCACAA GGATCACACC 3600

AGTTCOCCTG ATATAGGTGC AGAGGGCCCA GGTCTTCCCT CTAGCTAAGC CTTGGCCTTG 3660
 GCCTCCTACC CACACAGCAG CTGGTGCTT CCTGCCCTT GAGGCTAATA CATACTATGT 3720
 GGCCAGAAGA TGGTTTATGC TTTTAAAAA AATCTTATTT CAGAAATCTT TCCCTACTGT 3780
 TTTCTCCCA CATTATGTC TTAACAACACC TGTAGGGGAT TTTTTTTTTT TTTTTTTTTT 3840
 TGAGATGGAG TCTCGCTCTC GCCCAGGCTG GAGTGAATG GCGCGATCTT GGCTCACTGC 3900
 AAGGTCTGCC TCCCAGGTTT ACGCCATTCT CTGCTCCTAG CCTCCCCAGT AGCTGGGACT 3960
 ACAGGCGCCC GCTACCACGC CTGGCTAATT TTTTGCATT TTTAGTAGAG ACAGGGTTTC 4020
 ACTGTTAGCC AGGATGGTCT CGATCTCCTG ACCTCGTGAT CCACCCCTCT CAGCCTCCAA 4080
 AGTGCTGGGA TTAACAGGCA TGGAGCCCCA CCGCACTGGC CTGTATTGT GAGGAAGAAC 4140
 AGACCTCTT TAGAAGCCCT AGACTGCTGC CTCTGTAGT TCACTGGCAT CACTCAAAAT 4200
 ATTGTTGAG TTTCTTACTC ACTGAGTTGG TTTTATGTG TGGTGAAGG CGGAATCCT 4260
 CTTTTCATAT TCGTTCTCAT TGCCTATTGC TTTGCTCTAG TCCTATTACA ATCTTGTTTC 4320
 TTCCAG 4326

- 【0171】(2) 配列番号：13の情報： (ii) 分子の型：ゲノムDNA
 (i) 配列の特徴： (iii) 仮定：なし
 (A) 配列の長さ：166塩基対 (iv) アンチセンス：なし
 (B) 配列の型：核酸 (v) 断片の型：
 (C) 鎖の数：一本鎖 (vi) 起源：
 (D) トポロジー：直鎖状 (xi) 配列の記載：配列番号：13：

GAAGAGAGTT GTATGTACAA CCCAACAGGC AAGGCAGCTA AATGCAGAGG GTACAGAGAG 60
 ATCCCCGAGG GGAATGAGAA AGCCCTGAAG AGGGCAGTGG CCGGAGTGGG ACCTGTCTCT 120
 GTGGCCATTG ATGCAAGCCT GACCTCCTTC CAGTTTACA GCAAAG 166

- 【0172】(2) 配列番号：14の情報： (ii) 分子の型：ゲノムDNA
 (i) 配列の特徴： (iii) 仮定：なし
 (A) 配列の長さ：270塩基対 (iv) アンチセンス：なし
 (B) 配列の型：核酸 (v) 断片の型：
 (C) 鎖の数：一本鎖 (vi) 起源：
 (D) トポロジー：直鎖状 (xi) 配列の記載：配列番号：14：

GTAAGAAGCT GCTGATCCTA TACAGCACTG TCTTTTATGA TACAACTTG ATGGTTTCTC 60
 GAAGGACCTT GGGTATTTTC AGTACTTAGT TTTTGTATTC ACATGGAGGT GCCCAGAGAG 120
 AAATTAACAA CTGCTGCAGT ATGGAGCAGC ATCTCTGTGG TAAACCTCC TGACACGGAT 180
 GGAATTCTTC AAACAGTCTC CTAGACTGGG AGATCCACA GGGTGACCTT TGGATTGCAT 240
 AGAGCCTCAC GCTGGTAGTT TGTATTCTAG 270

- 【0173】(2) 配列番号：15の情報： (ii) 分子の型：ゲノムDNA
 (i) 配列の特徴： (iii) 仮定：なし
 (A) 配列の長さ：106塩基対 (iv) アンチセンス：なし
 (B) 配列の型：核酸 (v) 断片の型：
 (C) 鎖の数：一本鎖 (vi) 起源：
 (D) トポロジー：直鎖状 (xi) 配列の記載：配列番号：15：

GTGTGTATTA TGATGAAAGC TGCAATAGCG ATAATCTGAA CCATGCGGTT TTGGCAGTGG 60
 GATATGGAAT CCAGAAGGGA AACAAGCACT GGATAATTAA AAACAG 106

- 【0174】(2) 配列番号：16の情報： (ii) 分子の型：ゲノムDNA
 (i) 配列の特徴： (iii) 仮定：なし
 (A) 配列の長さ：2270塩基対 (iv) アンチセンス：なし
 (B) 配列の型：核酸 (v) 断片の型：
 (C) 鎖の数：一本鎖 (vi) 起源：
 (D) トポロジー：直鎖状 (xi) 配列の記載：配列番号：16：

GTAATGATGG GAACACTACT TTTGTTATTC AGTCACCTT TTAACACTCA ACCTCACCTC 60
 CAGCTTCCCG ATATTCTTT CTCTGTCCCA AATCAAGAAA AAATTATCTC AGAGTTCTCA 120
 CTCTATCTT CTCAGTCAGA GGCTCTTAAT TCTCAGTCTG AACTTAATG GCCAGTGTGT 180

TAGTCCATTT TGCATTGCCA CAAAAGAATA CCCGAGACTG GGTAGTTTAT AAAGAAACGA 240
 GGTTTGTTTG GCTATACAAA GCGTGGCACT AGTATCTGCT CAGCCTCTGA TGAGGCCTCA 300
 GAGCTTTTAC TCATGGCAGA AGGCAAAAGA GGGAGCAGGC ATGTCACATA GTGAGAGAGG 360
 GAGCAAGAGA GAGAGGGAGG TGCCGACTCT TTAAGAACC AGCTCTTGCA TGAACATAA 420
 GAGTGAGAAC TCACTCATCA CCAAGGCGAT GGCACCAAGC CATTCCATGA GGAATCCACT 480
 CTCATAACCC AAACACCTCC CACTATGCCC CACCTCCAC ATTGGGGATC ACATTTCAGC 540
 ATGAGACTGG GAGGGGACAC ACATCCAAAC CATATCCGCC AGACAATAGT GCTCAATTAT 600
 GTGCTGGGCA GATGCTCCCT GTGTGCAAGG TGCTTAGTGA CATACATAAA CCAACGAGCA 660
 GATGACACCT TCAGTGAGCT CAGAGCCCAA TAAGACAGAC CTAACAAACC ATGAGATAAA 720
 GCAGTACAAA GAACCAGCAG GAGCTTTGGA ATTACGTATT TTTACTTTCT TTTGTCTCTA 780
 ATGTGATCAG TTTCTTAGAT GGTTCOCATT AGCAATCTGT CTTTAACAGT AGGGGAGCAG 840
 CGTTAAAGGT TTAATATTCC TTTTGAACAG TTTTTTCTCT TCAAAATACA CTTAAGATAC 900
 ACGTATATAA GAACTTGCCA AAGATTGTGA AGAGAAACAT TTTTGTAGAA TAAGATATAA 960
 ACAAAAAAAG TTAGTGTTAC TTTCTATGT TGGGAACAA AGAAATCC AGGGTACCTT 1020
 GCTTCCCAT TCTCTTAGC ACCTTGTGAC TTTTGGGAG GGCAGATTG ATAACAATTA 1080
 TAGTTTTCT TCTCTGGCTG ATCACCATT ACCTGGCAGC AGCACTGGCT AAATCTCTG 1140
 TCCTTAGTGC CCTCAAGGA GCAGGAGCC TAGACTCTGG GTCGCTGACA GACTCACGCA 1200
 GTGGTGTGT TCAAACCTGA AGCAACTTTT TATATCAGG TTCCAACCA AGGTGAACCT 1260
 GAGCATCTTC CCAAGTCTCC CACAGCTTCT GTCCGTGTGT GTCCCTTCTC TTGACTOCCA 1320
 GGTCCAAGCA CTTACCTGT TCTTTCATGA TCAGGTACCA TGTGTGGAGA TAGCTTCAA 1380
 GAGAGCTGGG AGGAAGAAAG GACACACCCG GGCAGGATCA GGAACACTGG GGGCCCTGG 1440
 AGAAGGGGAG AGTGGGGGAG GGTACAGGTT TAAATAAAA TGTGTGGTA ATTAGAGAAT 1500
 TGCTGGTTGG GGAAAGAGGT CTGAAAACAA TTCAGGAAGA TAAACAAGAC AATCTCTCT 1560
 CTCTCTCTT TCTCAGTGG TCTCTCTGT CTCTAGTCT CGCTACTCAT TTCCTTAGTA 1620
 ATCTCATCCA CTCTCATAGT TTCATCCATC TCTCTATGG GGTTTACCC CAAATCAAGA 1680
 TCACCAGCTT CAGCCTCCTT CTATGCTCT AAACACAT TTTCAAGATT AATATTCCC 1740
 AAATACAGCT CTGATCATAT CACTCTCCCA CTCAAAATCC CTCAGTGGCT CCTCACGATG 1800
 ATGGGTCACA GAGTAAAGGT GAAGCTTTT AACCTTGCAG TAAAGGTAAT TCAACCTGAT 1860
 CTCAATCTGC CTTTCCAGAC ATCTCTCCA CTACACCTG TTAGGCACAC TGCTTTTCAG 1920
 CTACATGATC CTAACAGTGC CCCACATTT CTGCTCTG TTGTTCAATT CACACCTTC 1980
 CACTGGCATC CCCTTCCAC AGGTGAAAT TCTACTAGC CTTTGGCTC AGCTCAAATG 2040
 CCACTCTTA CATCAAGCCT CTAAGATTCT CTTGATCAGA AGGAATCTTT CCCTCCTTG 2100
 ATACCTACAG TATTATGCCT TCTCCTATT TCTTGACTTT AAACCTTTA AAGTTAAAA 2160
 ACATCATATT CATTTTGTG TACCATCAGT ACCTCGACA ATACTCAGTA AATATTTAA 2220
 TGAATAAATA AACTGAGAGT ACTAAGTATT TTTCTTGATT GGTCTTACAG 2270

【0175】(2) 配列番号: 17の情報: (ii) 分子の型: ゲノムDNA
 (i) 配列の特徴: (iii) 仮定: なし
 (A) 配列の長さ: 97塩基対 (iv) アンチセンス: なし
 (B) 配列の型: 核酸 (v) 断片の型:
 (C) 鎖の数: 一本鎖 (vi) 起源:
 (D) トポロジー: 直鎖状 (xi) 配列の記載: 配列番号: 17
 CTGGGGAGAA AACTGGGGAA ACAAAGGATA TATCCTCATG GCTCGAAATA AGAACAACGC 60
 CTGTGGCATT GCCAACCTGG CCAGCTTCCC CAAGATG 97

【0176】(2) 配列番号: 18の情報: (ii) 分子の型: ゲノムDNA
 (i) 配列の特徴: (iii) 仮定: なし
 (A) 配列の長さ: 598塩基対 (iv) アンチセンス: なし
 (B) 配列の型: 核酸 (v) 断片の型:
 (C) 鎖の数: 一本鎖 (vi) 起源:
 (D) トポロジー: 直鎖状 (xi) 配列の記載: 配列番号: 18:
 TGACTCCAGC CAGCCCAAAT CCATCCTGCT CTTCATTTTCTTCCACGAT GGTGCAGTGT 60

AACGATGCAC TTTGGAAGGG AGTTGGTGTG CTATTTTGA AGCAGATGTG GTGATACTGA 120
 GATTGTCTGT TCAGTTTCCC CATTGTGTTG TGCTTCAAAT GATCCTTCCT ACTTTGCTTC 180
 TCTCCACCCA TGACCTTTT CCACTGTGGC CATCAGGACT TTCCCCTGAC AGCTGTGTAC 240
 TCTTAGGCTA AGAGATGTGA CTACAGCCTG CCCCTGACTG TGTTGTCCCA GGGCTGATGC 300
 TGTACAGGTA CAGGCTGGAG ATTTTCACAT AGGTTAGATT CTCATTACAG GGAAGTAGTA 360
 GCTTTAAGCA CCCTAGAGGA CTAGGGTAAT CTGACTTCTC ACTTCCTAAG TTCCCTTCTA 420
 TATCCTCAAG GTAGAAATGT CTATGTTTTC TACTCCAATT CATAAATCTA TTCATAAGTC 480
 TTTGGTACAA GTTTACATGA TAAAAAGAAA TGTGATTTGT CTCCCTTCT TTGCACTTTT 540
 GAAATAAAGT ATTTATCTCC TGTCTACAGT TTAATAAATA GCATCTAGTA CACATTCA 598

- 【0177】 (2) 配列番号：19の情報： (ii) 分子の型：ゲノムDNA
 (i) 配列の特徴： (iii) 仮定：なし
 (A) 配列の長さ：459塩基対 (iv) アンチセンス：なし
 (B) 配列の型：核酸 (v) 断片の型：
 (C) 鎖の数：一本鎖 (vi) 起源：
 (D) トポロジー：直鎖状 (xi) 配列の記載：配列番号：19：

TTTTGTGTG GATACTGTGT TAGGTGCTGG AGGAAAAAG ATGAATAGAA CATCTTCTAT 60
 GTACTTCATG CGCTCACAGT CTGGTTGTAG AGACTGTCAC ATAAACATTT CATCCCAATT 120
 CATTATTTTG TTCATTCTT CAGCCAATAT ATATTGAGTT CTTACTCTGT GCCAAGAAGT 180
 GTACTACATT TCTGGGATTA AGTGGATATA AGGAGATCTC AGTGTTTAAT CTGCTGAGG 240
 GGAGACTAAA TTAAGTGACA TGGAACTTG GGTCTTGAAA AACATTTTAA GGTATTTT 300
 TCTTTTCTCT CTCTCTCGCT CTGTCTTTCT CTCTCTTTCTG TCAGGGTCTC CCTCTGTGTC 360
 CCAGGCTGGA GTCAGTGGCA CTCATAGCTC ACTGCAGCT TGATCTCCTG GGCTCAAGAG 420
 TTCTTCCAC CTCAGTCTCC TAAGTAGCTT GGACTACGG 459

- 【0178】 (2) 配列番号：20の情報： (ii) 分子の型：蛋白質
 (i) 配列の特徴： (iii) 仮定：なし
 (A) 配列の長さ：329アミノ酸 (iv) アンチセンス：なし
 (B) 配列の型：アミノ酸 (v) 断片の型：N-末端
 (C) 鎖の数：一本鎖 (vi) 起源：
 (D) トポロジー：直鎖状 (xi) 配列の記載：配列番号：20：

Met Trp Gly Leu Lys Val Leu Leu Leu Pro Val Val Ser Phe Ala Leu
 1 5 10 15
 Tyr Pro Glu Glu Ile Leu Asp Thr His Trp Glu Leu Trp Lys Lys Thr
 20 25 30
 His Arg Lys Gln Tyr Asn Asn Lys Val Asp Glu Ile Ser Arg Arg Leu
 35 40 45
 Ile Trp Glu Lys Asn Leu Lys Tyr Ile Ser Ile His Asn Leu Glu Ala
 50 55 60
 Ser Leu Gly Val His Thr Tyr Glu Leu Ala Met Asn His Leu Gly Asp
 65 70 75 80
 Met Thr Ser Glu Glu Val Val Gln Lys Met Thr Gly Leu Lys Val Pro
 85 90 95
 Leu Ser His Ser Arg Ser Asn Asp Thr Leu Tyr Ile Pro Glu Trp Glu
 100 105 110
 Gly Arg Ala Pro Asp Ser Val Asp Tyr Arg Lys Lys Gly Tyr Val Thr
 115 120 125
 Pro Val Lys Asn Gln Gly Gln Cys Gly Ser Cys Trp Ala Phe Ser Ser
 130 135 140
 Val Gly Ala Leu Glu Gly Gln Leu Lys Lys Lys Thr Gly Lys Leu Leu
 145 150 155 160
 Asn Leu Ser Pro Gln Asn Leu Val Asp Cys Val Ser Glu Asn Asp Gly

165 170 175
 Cys Gly Gly Gly Tyr Met Thr Asn Ala Phe Gln Tyr Val Gln Lys Asn
 180 185 190
 Arg Gly Ile Asp Ser Glu Asp Ala Tyr Pro Tyr Val Gly Gln Glu Glu
 195 200 205
 Ser Cys Met Tyr Asn Pro Thr Gly Lys Ala Ala Lys Cys Arg Gly Tyr
 210 215 220
 Arg Glu Ile Pro Glu Gly Asn Glu Lys Ala Leu Lys Arg Ala Val Ala
 225 230 235 240
 Arg Val Gly Pro Val Ser Val Ala Ile Asp Ala Ser Leu Thr Ser Phe
 245 250 255
 Gln Phe Tyr Ser Lys Gly Val Tyr Tyr Asp Glu Ser Cys Asn Ser Asp
 260 265 270
 Asn Leu Asn His Ala Val Leu Ala Val Gly Tyr Gly Ile Gln Lys Gly
 275 280 285
 Asn Lys His Trp Ile Ile Lys Asn Ser Trp Gly Glu Asn Trp Gly Asn
 290 295 300
 Lys Gly Tyr Ile Leu Met Ala Arg Asn Lys Asn Asn Ala Cys Gly Ile
 305 310 315 320
 Ala Asn Leu Ala Ser Phe Pro Lys Met
 325

【0179】(2) 配列番号：21の情報：

(i) 配列の特徴：

(A) 配列の長さ：20塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

CCGAAACGAA GCCAGACAAC

【0180】(2) 配列番号：22の情報：

(i) 配列の特徴：

(A) 配列の長さ：21塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

GCTCACCACA GGTAGCAGCA G

【0181】(2) 配列番号：23の情報：

(i) 配列の特徴：

(A) 配列の長さ：21塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

CTGCTGCTAC CTGTGGTGAG C

【0182】(2) 配列番号：24の情報：

(i) 配列の特徴：

(A) 配列の長さ：20塩基対

(B) 配列の型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

CCCAAATTAA ACGCCGAGAG

【0183】(2) 配列番号：25の情報：

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：21：

20

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：22：

21

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：23：

21

(ii) 分子の型：ゲノムDNA

(iii) 仮定：なし

(iv) アンチセンス：なし

(v) 断片の型：

(vi) 起源：

(xi) 配列の記載：配列番号：24：

20

(i) 配列の特徴：

- (A) 配列の長さ：20塩基対
- (B) 配列の型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状
- (ii) 分子の型：ゲノムDNA

CTCTCGGCGT TTAATTTGGG

【0184】(2) 配列番号：26の情報：

- (i) 配列の特徴：
- (A) 配列の長さ：21塩基対
- (B) 配列の型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状

GGTACTTTGA GTCCAGTCAT C

【0185】(2) 配列番号：27の情報：

- (i) 配列の特徴：
- (A) 配列の長さ：20塩基対
- (B) 配列の型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状

CCGACTCTG TCGACTATCG

【0186】(2) 配列番号：28の情報：

- (i) 配列の特徴：
- (A) 配列の長さ：21塩基対
- (B) 配列の型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状

CACATATGGG TAGGCATCTT C

【0187】(2) 配列番号：29の情報：

- (i) 配列の特徴：
- (A) 配列の長さ：21塩基対
- (B) 配列の型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状

GAAGATGCCT ACCCATATGT G

【0188】(2) 配列番号：30の情報：

- (i) 配列の特徴：
- (A) 配列の長さ：21塩基対
- (B) 配列の型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状

GTTACATTAT CGCTATTGCA C

【0189】(2) 配列番号：31の情報：

- (i) 配列の特徴：
- (A) 配列の長さ：21塩基対
- (B) 配列の型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状

GCAAAGGTGT GTATTATGAT G

【0190】(2) 配列番号：32の情報：

- (i) 配列の特徴：

- (iii) 仮定：なし
- (iv) アンチセンス：なし
- (v) 断片の型：
- (vi) 起源：
- (xi) 配列の記載：配列番号：25：20

- (ii) 分子の型：ゲノムDNA
- (iii) 仮定：なし
- (iv) アンチセンス：なし
- (v) 断片の型：
- (vi) 起源：
- (xi) 配列の記載：配列番号：26：21

- (ii) 分子の型：ゲノムDNA
- (iii) 仮定：なし
- (iv) アンチセンス：なし
- (v) 断片の型：
- (vi) 起源：
- (xi) 配列の記載：配列番号：27：20

- (ii) 分子の型：ゲノムDNA
- (iii) 仮定：なし
- (iv) アンチセンス：なし
- (v) 断片の型：
- (vi) 起源：
- (xi) 配列の記載：配列番号：28：21

- (ii) 分子の型：ゲノムDNA
- (iii) 仮定：なし
- (iv) アンチセンス：なし
- (v) 断片の型：
- (vi) 起源：
- (xi) 配列の記載：配列番号：29：21

- (ii) 分子の型：ゲノムDNA
- (iii) 仮定：なし
- (iv) アンチセンス：なし
- (v) 断片の型：
- (vi) 起源：
- (xi) 配列の記載：配列番号：30：21

- (ii) 分子の型：ゲノムDNA
- (iii) 仮定：なし
- (iv) アンチセンス：なし
- (v) 断片の型：
- (vi) 起源：
- (xi) 配列の記載：配列番号：31：21

- (A) 配列の長さ：21塩基対
- (B) 配列の型：核酸

- (C) 鎖の数：一本鎖
(D) トポロジー：直鎖状
(ii) 分子の型：ゲノムDNA
(iii) 仮定：なし

GCCGTTGTTT TTATTTGAG C

【図面の簡単な説明】

【図1】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図2】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図3】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図4】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図5】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図6】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図7】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図8】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図9】 ヒトカテプシンKのゲノムヌクレオチド配列
【配列番号：1】を示す。エキソンを大文字で示し、下線を引く。イントロンを小文字で示す。

【図10】 ヒトカテプシンKのヌクレオチド、エキソン-イントロン境界および推定アミノ酸配列を示す。

【図11】 ヒトカテプシンKのヌクレオチド、エキソン-イントロン境界および推定アミノ酸配列を示す。

【図12】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図13】 カテプシンK【配列番号：2-19】の構

- (iv) アンチセンス：なし
(v) 断片の型：
(vi) 起源：
(xi) 配列の記載：配列番号：32：

21

造的特徴を示す。

【図14】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図15】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図16】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図17】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図18】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図19】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図20】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図21】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図22】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図23】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図24】 カテプシンK【配列番号：2-19】の構造的特徴を示す。

【図25】 イントロン-エキソン接合を示す

【図26】 カテプシンK、ヒトカテプシンS、L、H、B、D、E、GならびにウサギOC2ポリペプチドのアミノ酸配列の間の類似領域を示す。

【図27】 カテプシンK、ヒトカテプシンS、L、H、B、D、E、GならびにウサギOC2ポリペプチドのアミノ酸配列の間の類似領域を示す。

【図28】 ヒトカテプシンKの推定アミノ酸配列を示す。

【図29】 ゲノムDNAを増幅するためのPCRプライマーを示す。

[SEQ ID NO. 1]

GCTTTGGCTCCCAAAGGCCTGGGATTACAGGCGTGAACCACTGCGCCTAG
CCTGTTAGCAGCTCTTAAAAATCCAGAGGCATAAGCCTGTATTTTGGAGGG
TTTATGCATGGAATCCAGCTAGAACTGAGTCTATTACAGATCCCATTTA
TTATCCTTTCTATTCCAAGAAGCCTTTTCTCCTTCCCCACATCTGTT
TATGGAAGAAAATGAAGTTTGGGGTGTGTTTGAGGAATCAGCTAGATTC
TTATGATCTGTCACATGCTTGGATGTTGGGGAAGCATTGGAGAGCTCA
TGTGACTTGTCTAGATTGGGGATTTAATTGAGACAGATGATGTTATC
GGGCATCCACCACCTGAGAGTTTACCAACAGAGTCACATGTGAGTCCA
TCAGAACTTACGGCATTGATTCAAGTGTGTCATAAATAACCAGGACTGC
TGTTTTGGTTACTTTTAAAGACAGTTTCATCTGGACTTTCTGGGCATAT
CCTCCTTCAGCAAAACCACATTAGGCTGGGAAAATATTCTGCCTGGAAG
TAATGACAACTTGCAACCAACAGCTTATAAAAATACAAAGAATTCTGGA
GCCTATGGCTTCCATTACATTATTCTTTATAGCCTTTTATGTTTATTAC
CGCATCCCAGAGGTGAGAGTCAGACACAAATATGAAAATAGGTTTCAATG
TTGGAGAGGTAAATCCTAACAGGAAAGGGGTAGGAAAAGATATAATCCCC
CAATATTAATAAAGATATTGAAGAAGAAGGATGGGAGAGACTAGGGCT
GTGTCCTTCTTTTACTACCAAAAGAGAAAGTAAGCTCCTATTGAGTC
AATAGATATTGAGGTCTTGTATTGTCACCAAGACAGTCTTGTGAGAC
TAAATAGCTAGTAATCCCTACCCTGGCACACATGCTGCATACACACAGA
AACACTGCAATCCACTGCCTCCTTCCCTCCTCCTACCCTTCTTCTCT
CAGCATTCTATCCCCGCTCCTCCTTACCCAAATTTCCAGCCGATC
ACTGGAGCTGACTTCCGCAATCCCGATGGAATAAATCTAGCACCCCTGAT
GGTGTGCCCAGACTTTGCTGCCGAAACGAAGCCAGACAACAGATTTCAT
CAGCAGgtaacgtttg caacttccta gatcttttag cttttcattcc
tgtcaattctctgagtattagggatgtagtgacttgaggatcacataaa
cttttagcctctgcagatgaaaacagagatgcacttcttaggtcattccc
tggctaaataaaatctgcctggaaatctgtagaattccttgtatgattta
tatatatacatatcatgattgttagtaaaagcaaagtatatagggaaatcat
tccccatccttcaagagtggccttctgcagtgttttctactttggcca
acaaggatcaaaaacgggttaactccttagtgaggaggaggagagtggtatg
gggaggtagtagctcagtgcttctgttccactgagacatctcaaagccct
taacactctagtttttaaatgtcctactggacattttgccagtttgcaaa
attacatgtaaatggactataagcaattgtgtaagccatatgtcatgtgtg
caggctgcaaatgttctttaaattggaggatttgttaattaagaagccaa
tgcaagaaatgagtgaagctaactagagtaaaacttatgaaaagctgtgaa
tttcatcatcatagaaacattgcttttcagtctgaacattcttctaacaaa
ccttggatctgaggcttcttgtcctttgcggcagccacagtgggtttttg
ttgttaggggaaaaataaaaaaccttgccgcagcatctggttaagattag
ggcagtttccctgcctaaggagggaagggagagaaaaaggaagaagaatg
cataaggagaaatgaggagatatataatgtctcagaaaacaggaaacattg
tcctattttcccttgtcctcttctgacaagatctgggaaagtaccagaat
ttaggcacgaaagagaagaacgcctcgaagaaatgatcaggaagcaaaac
ttagacggaaatctctcctttgtgtattctgaacccactaccaccttgc
tatttgtctg tctccaagcc tgctagggac cctggaggaa acgcactgag
cccattctga ttgtccagtt tctatcccc atttctggtt gtgtacgtgt
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gagagagaga

【図2】

(図1のフズキ)

gagacagaga gagaaacaga gagagtgtgt gttgcctaaa tctcccgaga
gagagagaga gagagagaga gagagagaga gagagaaaag agagaaatgg
ctaaatcccc ctagatcaaa gtccttggaa ccagatgtac cagcatccta
tctaaacaca ggccccctct gactatcatt gttttatcac cctttttccg
tctacctttc tcttcctcat aaagcctagt tttcctctgt ttccttgcca
aatggaagag ttttccctaa ctacattctt ctgcagGATG TGGGGGCTCA
AGGTTCTGCT GCTACCTGTG GTGAGCTTTG CTCTGTACCC TGAGGAGATA
CTGGACACCC ACTGGGAGCT ATGGAAGAAG ACCCACAGGA AGCAATATAA
CAACAAGgtg cctgggggtcc tggagggggc atggcaggaa ggctgag
acctgagctc tctcatctta gcttcagac tcccttcttc aatccaaatg
ctttattcca agcaaatacag tccctcttcc ctaactcatg ttaacatacg
gttttcatto ctatgcttca atcatcctct tgtcaaactt gtattccttc
ccttttggttt tataagtgtg taacattcct cttttgggaa gagtcccaag
attaatgctg ttaatccata agcaattttt ctgtctctcc agagcttgtg
tggttgttta catattatct ctcttcttgc aggtctctaa ttocatggtt
agttcccca ctaaactgta aacttttatg attgtgagtt tcccttattc
tcttaaaacc cttcacaata ttacatatga actgtagaca gtctatacaa
gtactgactatgctttgttttagGTGGATGAAATCTCTCGGCGTTTAATTTGGGA
AAAAAACCTGAAGTATATTTCCATCCATAACCTTGAGGCTTCTCTTGGTGTCCA
TACATATGAACCTGGCTATGAACCACCTGGGGGACATGgcaagtatagcttcagc
tccgtgcccacctgcaccatttgccttagttccctgctgatgcctggcctcttt
cttctttgtcttagACCAGTGAAGAGGTGGTTCAGAAATGACTGGACTCAAAGTA
CCCCTGTCTCATTCGCCGAGTAATGACACCCTTTATATCCCAGAATGGGAAGGTAG
AGCCCCAGACTCTGTCTGACTATCGAAAGAAAGGATATGTTACTCCTGTCAAAAATC
AGgt actctccttt cttctgggtg tgcataatgta atctggca
tgaccttttc ctttttctgc tgctttgttc ttgaggtgaa agggcaccag
gaaaagaggg caaggaatta aggtacatct ccccatctcc attctgttat
ttaacctcat ttgtttctgt acatttgggt tgbbctctgt ttttctttt
cttttccctt tttttttttt tttttttttt gagatagagt ctactctgt
cgcccaggat ggagtgcagt ggtgcaatct tggctcactg caacctacac
ctcccgggtt caagcgattc tctgcctca gcctcctgag tagctgagat

【図3】

(図2 の つづき)

tacaggcacg	cgccactacg	cctgggcta	ttttctat	ttatagagat
gcgttttcac	catgttggcc	aggctgggt	tgaactgacc	tcagggtgatc
cacctgcctc	agcctcccaa	agtgtggga	ttagagtc	gagccatcgc
ggcctgggtt	ttctttatta	caaatagt	tgcaataagc	acccttgtgc
atatgttttt	gtgcacatgt	acaaata	atgcaaaata	agtcctaaaa
ttggaattgt	taggtcaca	ataatcctt	ccccccccc	aaattttttt
ttttttttt	agacagcgtc	tctgtcacc	aggctggagt	ccagtggcgc
aatcatgggt	cactgcagcc	tcaacgtct	aggctcaagt	gattctccaa
cctcagcctc	cctagttagt	gggaattaga	agcacatg	accacaccca
gctaatttta	aaaaattttt	tgtagagac	aggtttttgc	catgctaccc
aagctgggt	caaattcctg	ggctcaagca	atctgcccgc	ttcggcctcc
caaagtgt	ggattacaga	catgagccac	catgcccagc	ccaaaaaagt
ttttgcaatc	ttacattctt	actagcatga	gaatgtcagt	tttttcacaa
cccaaacaac	acaggattgt	atcagcaaga	taaacaattg	atttaacggt
catttaacaa	acactttttg	accccccaga	cctaccagat	gcagtgttag
gcagcagaga	ctcaagatga	ctaagacaca	acctgtgtcc	tcaggaaatc
tcaatctaa	aaaatagaac	eggaaagaaa	gaaaaatcta	caatctagct
gcacaaacaa	taatagctaa	tactttttga	gattttattg	tttgtcagga
acttcttaac	tctttacatg	agtttaaata	tttaatccct	tataacaata
ttttatgcat	agagaaactg	agacacaggc	aaatttagta	acttaccogg
ggtcacatag	ctactgggtg	gcaaagtcag	ggttagctcc	caggacaaat
gcctccacag	ctgggtactgt	gctctgcttt	actgtagcta	atagtaaaaa
tggtagcaaa	aatcaatagc	agtagaacag	tgcaacagat	attaagcgga
agaggaagac	tcacaacaat	gacaacattt	gtgctgaaat	ttttaagaac
acatggaatt	tccttcagcc	gggtagagag	aagatataga	aatgtaaaaca
ccaaagattc	atagtttctc	tgtatccctt	tcagGGTCAGTGTGGTTTCCTG	
TTGGGCTTTTAGCTCTGTGGGTGCCCTGGAGGGCCAACTCAAGAAGAAA	ACTGG			
CAA	ACTCTTAAATCTGAGTCCCCAGAACCTAGTGGATTGTGTGTCTGAGAATGA			

(図3のつづき)

TGGCTGTGGAGGGGGCTACATGACCAATGCCTTCCAATATGTGCAGAAGAACCGGG
GTATTGACTCTGAAGATGCCTACCCATATGTGGGACAGtg agattgctcc
 acacaattat acagctctgt tggctcctcc ttccccagca tgatggt
 ttgtactgga aacaattcca gaaatactgt tttctgttat cctatcctgc
 tttcttgatg gaataatttc ccacagaagg ccaagaagat ttccacaatc
 tgggggaatt tagggagctt aagctactat agctcctatt tgcattctctg
 ccatggagag aaaacagagg ctaggctacc taccatag acttccgagc
 tgggttctat aaccctctgc tcaattcctc actcccacaa caaaccacaa
 aaccaccat gctattttca caaattgtgt ggctttatct tatatgatct
 cagtgtgagt tttcagaaca tttcagcaaa ttatgtaagt ttacatgcta
 acatctataa aatgagagaa aaaacaagtt gcttcatata agagataagg
 gattaactca gttcctcctg catgatcctc tagtcatagg aaggaaatca
 tatctgaaag ggaggcaacc tgaggggttt tttatacaca tagggctggg
 tctgatagac aatataatgt agggccttca caacagaaac ctctgaaaca
 gggacagcaa gtttgagaat aaaaatgatg gctactgtgt tctaagccgt
 gtccttagtg cattttttct ttttcttttt ttcatttaac ctcataacaa
 ctctgttagg tagacttatc ttgaatgtat aggtgaggaa atggacactt
 aaggagataa gacagtataa ttcataccac tagtatgtaa caatgtaaga
 tgtatctacc agggatgttt atcttctgca aacattccta ggtatatctc
 ccatgcacat gtgcaagaat ttcttactag gatataatgc cttggaactg
 aattgtctgg gtcttagggg atgtctgtct tcaactttac tacacaatgt
 caaattgttt gccaaaatat ttggaaaaat ttatacctgc aatgtgtaag
 aaatcccctt caatcacctt tttatcagta tgtttatctg gccatttgca
 tttcttcttc agtgaattaa ctgtttttat ctcttgctca tttgtttttc
 tttttatctt tttgaaatag ggtcttactc tgttgcccaa ggctggagtg
 tggtgataca gtcatagctc actgcagcct ccacttccgg gctcaagcaa
 tcctctcgcc tcagcctccc aaatagctag gatataggtg catgccatca
 tgcccaccaa tttcaaaaaa cttttgaaat tttttttttg taaaagctag
 gcatgggtggc tcatgcctgt aatcccagca ctttgggagg ctgaggtggg

(図4のつぎ)

```

aggatcgctt gagcccagga attggaggtc ggcctgatac aacatagcaa
gacctcatct ctacagaaaa aattttttaa agtagccagg tatgatggcg
tgcatagttc tagctactcc ggaagctggg tgggaggaca acttgagcct
gggagttcaa ggctgctgtg aactgtgata atgtcactgc tctctagcct
gggtgacaga gtgagaccct gtcccccacaa acaacaaccg ttttttttgg
tagagacatt gtctcgctat gttgccaagg ctagtctcaa actcctgggc
tcaagcaatc ctcccacctc ccaaagtgtt gggattatag atgtaagcca
ccatgcctgg cctacccttt tttttttttt ttgaaatgga agttttgctt
ttgtcaccta ggcttgagtg cagtggcgcg atcttggttc actgcaacct
ccacctcttg gattcaagca attctctctg ctcagcctcc tgaatagctg
ggattatagg caccgcgaac cacgcccggc tagtttttgt attttttagta
cagacagggg ttcaccatgt tggccagctg gtcttgaacc cctgacctca
gggtggctccgc ccgcctcggc ctcccaaaat gctgggatta aaagtgtgag
ccaccatgcc ccaccctta ctcatTTTTT attggattgt ttttctctt
tcttagcgat tcttaaaagt ttaaagagaa tatttggata caatactatg
tattttaaag ttgaggctctg tctttocatt cttctcacga tgtctttcaa
tctagaaaag ttaattttaa taggcctggc gcggtggctc acgcttgtaa
tcccagcact ttgggaggct gagatgggtg gatcacaagg tcaggagatg
aagaccatcc tggctaacat gggtgaaacc ctgtttctac taaaaatata
aaaaaattag ctgggcgtgg tggcaggctg ctgtagttcc agctactcgg
gaggctgagg caggagaatg gcgtgaaccc gggagggtga gcttgagctg
agccgagatt gcaccactgc actccagcct gggcaactga gcaagactgc
gtttcaaaaa aaaaaaaagt taattttaat atagtataat tagttaaagg
attaattttc cctttgcaat ttttgtaatg tgttttattc gtttatgaat
ggagaaaggt aagaaaaaat aaaattttaa aaagaagaga tgtggccagg
tacgggtggc cacacctata atcccagtag tttgggaggc tgaggcaggc
agatcacttg aggtcaggag tttgagacca gctgggataa catggtgaaa
ccccatctct actaaaaata caaaaattag ccagggtgtga ttgcgcacgc
ttgtaatccc agcaggctga ggcaggagaa ttgctcgaa tcaggaggca

```

(図5のつづき)

gaggttgcag tgagccaaga tcatgccatt gcactccagc ctgggtaaca
gagactctgt ttcaaaaaaa taaaaagata aaaaaggaag agatctgata
gggcgccag ataaacattt taaaggggat ggtattataa gtttgttccc
agcataatgc caggttattc tgactttaaa gtatcatcac ataatatctt
tttgagtcaa tttccaagat attctgtttc acttgtaatt ctgtgtaatt
tttggcacca ggaggcatca gggatttggg gcacatggca gaaacaaagg
catcttgaaa aatatcaagg cagtagacca ctgtaatctt aaaatggcat
atcaaagtct gctattgctg ttaatattta gataatgtta gataatgtat
tttttttagag ggtatctcac tatcttgcac aggctggagt agagtggcta
ttcacagcat gatcacagta cactaaaggc tcaaactcct gggcacaaac
aatcctcctg cctcagcctg ctgagtagta gataataagt tcttgtggat
gcaaccttag ggttctgaag gggtagtctg taggaaaatg aattgctgaa
aagaatacac caccttaaca tgggctatta ttcgattcca taattgtggc
ttgccaatga aacattgcta actacctgta aaatatagtg ttggaagtca
taggctaaat tgctaagttc tttaatctat tttagtgtct tgttatgtac
ttttatatatt tgtctttgat gagagcacia ggatcacacc agttcccctg
atataggtgc agagggccca ggtcttcctt ctgctaagc cttggccttg
gcctcctacc cacacagcag ctggtgcctt cctgccccct gaggctaata
catactatgt ggccagaaga tggtttatgc tttttaaaaa aatcttattt
cagaaatctt tccctactgt tttcctccca catttatgtc ttaaaacacc
tgtaggggat tttttttttt tttttttttt tgagatggag tctcgctctc
gcccaggctg gagtgcaatg gcgcgatctt ggctcactgc aaggctctgc
tcccaggttc acgccattct cctgcctcag cctccccagt agctgggact
acaggcgccc gctaccacgc ctggctaatt tttttgcatt tttagtagag
acagggtttc actgttagcc aggatggtct cgatctcctg acctcgtgat
ccaccctctt cagcctccaa agtgctggga ttaacaggca tggagcccca
ccgcactggc ctgtatttgt gaggaagaac agaccctctt tagaagccct
agactgctgc ctctgttagt tcaactggcat cactcaaaat attgggttag
tttcttactc actgagttgg tttttatgtg tggtaggaagg cgggaatcct

(図6のつづき)

cttttccatat tcgtttctcat tgcctattgc tttgtcctag tcctattaca
 atcttgttttcttccagGAAGAGAGTTGTATGTACAACCCAACAGGCAAGGCAGCTA
 AATGCAGAGGGTACAGAGAGATCCCCGAGGGGAATGAGAAAGCCCTGAAGAGGGCA
 GTGGCCCGAGTGGGACCTGTCTCTGTGGCCATTGATGCAAGCCTGACCTC
 CTTCAGTTTTACAGCAAAG gtaagaagct gctgaccta tacagcactg
 tcttttatga tacaacttg atggtttctc gaaggacctt gggtattttc
 agtacttagt ttttgtattc acatggaggt ggccagagag aaatgaacaa
 ctgctgcagt atggagcagc atctctgtgg taaacctcc tgacacggat
 ggaattcttc aaacagtctc ctagactggg agatcccaca gggtgacctt
 tggattgcat agagcctcac gctggtagtt tgtattctag GTGTGTATTA
 TGATGAAAGCTGCAATAGCGATAATCTGAACCATGCGGTTTTGGCAGTGGGAT
 ATGGAATCCA GAAGGGAAACAAGCACTGGATAATTAAAAACAGgtaatgatg
 ggaacactac ttttgttatt cagtcacctt ttaaacactc aacctcacct
 ccagcttccc gatattcctt tctctgtccc aaatcaagaa aaaattatct
 cagagttctc acttctatct tctcagtcag aggtcttta tctcagttct
 gacacttaat ggccagtgtg ttagtccatt ttgcattgcc acaaaagaat
 acccgagact gggtagttta taaagaaacg aggtttgttt ggctatacaa
 agcgtggcac tagtatctgc tcagcctctg atgaggcctc agagctttta
 ctcattggcag aaggcaaaag agggagcagg catgtcacat agtgagagag
 ggagcaagag agagagggag gtgcccactc tttaaagaac cagctcttgc
 atgaactaat agagtgagaa ctactcatc accaaggcga tggcaccaag
 ccattccatg aggaatccac tctcataacc caaacacctc ccactatgcc
 ccacctccca cattggggat cacatttcag catgagactg ggaggggaca
 cacatccaaa ccatatccgc cagacaatag tgctcaatta tgtgctgggc
 agatgctccc tgtgtgcaag gtgcttagtg acatacataa accaacgagc
 agatgacacc ttcagtgagc tcagagccca ataagacaga cctaactaac
 catgagataa agcagtacaa agaaccagca ggagctttgg aattacgtat
 ttttactttc ttttgtctct aatgtgatca gtttcttaga tggtttccat
 tagcaatctg tctttaacag taggggagca gcgttaaagg tttaatattc

【図8】

(127の77"芝)

cttttgaaca gtttttttcc ttcaaaatac acttaagata cacgtatata
 agaacttgcc aaagattgtg aagagaaaca ttttttagaa ataagatata
 aacaaaaaaa gttagtgtta ctttcctatg ttggggaaca aagaaaactc
 caggggtacct tgcttcccat ttctcttttag caccttgtga cttttgggga
 ggggcagatt gataacaatt atagttttcc tttcctggct gatcaccatt
 aacctggcag cagcactggc taaatctcct gtcccttagtg cctccaagg
 agcaggagcc ctagactctg ggtcgctgac agactcacgc agtggtgttg
 ttcaaacctg aagcaacttt ttatatcaca gttccaactc aagggtgaacc
 tgagcatctt cccaagtctc ccacagcttc tgtcctgtgt tgtcccttct
 cttgactccc aggtccaagc acttaccctg tttttctatg atcagggtacc
 atgtgtggag atagcttcca agagagctgg gaggaagaaa ggacacacccc
 gggcaggatc aggaacactg gggggccctg gagaagggga gagggtggga
 ggggtacaggt tttaaataaa atgtgttggg aattagagaa ttgctgggtg
 gggaaagagg tctgaaaaca attcaggaag ataaacaaga caatctctcc
 tctctcctct ttctcacgct gtctctcttg tcttctagtc tctgactca
 tttccttagt aatctcatcc actctcatag tttcatccat ctctcctatg
 ggggtttacc ccaaatcaag atcaccagct tcagcctcct tcttatgctc
 taaactcaca ttttcaagat taatattccc caaatcacgc tctgatcata
 tcactctccc actcaaaatc cctcactggc tcctcacgat gatgggtcac
 agagtaaagg tgaagctttt taaccttgca gtaaaggtaa ttcaacctga
 tctcaatctg cctttccaga catctctccc actacaccct gttaggcaca
 ctgcttttca gctacatgat cctaacagtg cccacactt tcctgcctct
 gttgttcatt tcacacccct ccactggcat ccccttccca cagggtcgaaa
 ttctacttag ccttttggct cagctcaaat gccacctctt acatcaagcc
 tctaagattc tcttgatcag aaggaaatctt tccctccttt gatacctaca
 gtattatgcc ttctccctat ttcttgactt taaactcttt aaagttaaaa
 aacatcatat tcatttttgt gtaccatcag tacctcgac aatactcagt
 aaatatitita atgaataaat aaactgagag tactaagtat ttttcttgat
 tgggtcttacagCTGGGGGAGAAAACCTGGGGAAACAAAGGATATATCCTCATGGCT

(12) 8のつづき)

CGAAATAAGAACAAACGCCTGTGGCATTGCCAACC'TGGCCAGCTTCCCCAAGATGTG
 ACTCCAGCCAGCCCAAATCCATCCTGCTCTTCCATTTCCTTCCACGATGGTGCAGT
 GTAACGATGCAC'TTTGGAAGGGAGTTGGTGTGCTATTTTGAAGCAGATG
 TGGTGATACTGAGATTGTCTGTTCAAGTTTCCCCATTGTGTTGTGCTTCAAATGA
 TCCTTCCTACTTTGCTTCTCTCCACCCATGACCTTTTCCACTGTGGCCATCAGGA
 CTTTCCCCCTGACAGCTGTGTACTCTTAGGCTAAGAGATGTGACTACAGCCTGCCCC
 TGACTGTGTTGTCCAGGGCTGATGCTGTACAGGTACAGGCTGGAGATTTTCACAT
 AGGTTAGATTCTCATTCAAGGGACTAGTTAGCTTTAAGCACCCCTAGAGGACTAGGG
 TAATCTGACTTCTCACTTCCTAAGTTCGCTTCTATATCCTCAAGGTAGAAATGTCT
 ATGTTTCTACTCCAATTCATAAATCTATTTCATAAGTCTTTGGTACAAAGTTTACAT
 GATAAAAAGAAATGTGATTTGTCTTCCCTTCTTTGCAC'TTTTGAATAAAGTATTT
 ATCTCCTGTCTACAGTTTAATAAATAGCATCTAGTACACATTCA'TTTTGTGTTGGA
 TACTGTGTTAGGTGCTGGAGGAAAAAGATGAATAGAACATCTTCTATGTACTTGA
 TGGCCTCACAGTCTGGTTGTAGAGACTGTCACATAAACATTTTCATCCCAATTCATT
 TATTTGTTCA'TTCCCTCAGCCAATATATATTGAGTTCTTACTCTGTGCCAAGAACT
 GTACTACATTTCTGGGATTAAGTGGATATAAG
 GAGATCTCAGTGT'TTAATCTGCCTGAGGGGAGACTAAATTAAGTGACATGGAACT
 TGGGTCTTGAAAAACATTTTAAGGTTATTTTCTTTTCTCTCTCTCTCGCT
 CTGTCTTTCTC TCTCTTTCGTGAGGGTCTCCCTCTGTTGCCCAGGCTGGAGTC
 AGTGGCACTCATAGCTCACTGCAGCCTTGATCTCCTGGGCTCAAGAGTTCTTCCCA
 CCTCAGTCTCCTAAGTAGCTTGGACTACGG

[図10]

cDNA CACACTTTGCTGCCGAACGAAGCCAGACACAGATTTCATCAGCAG¹G 49
 18
 ATGTGGGGGCTCAAGGTTCTGCTGCTACCTGTGGTGAGCTTTGCTCTGTACCCCTGAGGAG 109
 M W G L K V L L L P V V S F A L Y P E E
 22F 18 17
 ATACTGGACCCCACTGGGAGCTATGGAAGAAGACCCACAGGAAGCAATATAACAACAG¹ 169
 I L D T H W E L W K K T H R K Q Y N N K
 22R 3F
 GTGGATGAAATCTCTCGGCGTTTAAATTTGGGAAAAAACCTGAAGTATATTTCCATCCAT 229
 V D E I S R R L I W E K N L K Y I S I H
 289
 AACCTTGAGGCTTCTCTGGTGTCCATACATATGAAGTGGCTATGAACCACCTGGGGGAC 289
 N L E A S L G V H T Y E L A M N H L G D
 17
 ATG¹ACCAGTGAAGAGGTGGTTTCAAGATGACTGGACTCAAAGTACCCCTGTCTCATTC 349
 M T S E E V V Q K M T G L K V P L S H S
 3R
 CGCAGTAATGACACCCCTTATATCCAGATGGGAAGGTAGAGCCCCAGACTCTGTCTGAC 409
 R S N D T L Y I P E W E G R A P D S V D
 17
 TATCGAAGAAAGGATATGTTACTCTCTCAAAAATCAG¹GCTCAGTGTGGTTCCTGTGG 469
 Y R K K G Y V T P V K N Q G Q C G S C W
 4F
 GCTTTTAGCTCTGTGGTGGCCCTGGAGGGCCAACTCAAGAAGAAAAGTGGCAAACTCTTA 529
 A F S S V G A L E G Q L K K K T G K L L
 589
 AATCTGAGTCCCCAGAACCTAGTGGATTGTGTCTGTGAGAATGATGGCTGTGGAGGGGGC 589
 N L S P Q N L V D C V S E N D G C G G G
 649
 TACATGACCAATGCCCTTCCAATATGTGCAGAAGAACCAGGGGTATTGACTCTGAAGATGCC 649
 Y M T N A F Q Y V Q K N R G I D S E D A
 17 4R
 TACCCATATGTGGGACAG¹GAAGAGAGTTGTATGTACAACCCCAAGGCAAGGCAGCTAAA 709
 Y P Y V G Q E E S C M Y N P T G K A A K
 546F
 TGCAGAGGGTACAGAGAGATCCCCGAGGGGATGAGAAAGCCCTGAAGAGGGCAGTGGCC 769
 C R G Y R E I P E G N E K A L K R A V A
 829
 CGAGTGGGACCTCTCTGTGGCCATTGATGCAAGCCTGACCTCCTTCCAGTTTACAGC 829
 R V G P V S V A I D A S L T S F Q F Y S
 17
 AAAG¹GTGTGTATTATGATGAAAGCTGCAATAGCGATAATCTGAACCATGCCGTTTGGCA 889
 K G V Y Y D E S C N S D N L N H A V L A
 7F 546R 17

(12) 10の77"3)

GTGGGATATGGATCCAGAAGGGAAACAAGCACTGGATAATTAAAAACAG¹CTGGGGAGAA 949
 V G Y G I Q K G N K H W I I K N S W G E

 AACTGGGGAAACAAAGGATATATCCTCATGGCTCGAAATAAGAACAAAGCCCTGTGGCATT 1009
 N W G N K G Y I L M A R N K N N A C G I
 7R←—————8F
 GCCAACCTGGCCAGCTTCCCCAAGATGTGACTCCAGCCAGCCCAARTCCATCCTGCTCTT 1069
 A N L A S F P K M
 CCATTTCTTCCACGATGGTGCAGTGTAAAGGATGCACCTTTGGAAGGGAGTTGGTGTGCTA 1129

 TTTTGAAGCAGATGTGCTGATACTGAGATTGTCTGTTCAGTTTCCCCATTGTGTTGTGC 1189

 TTCAATGATCCTTCCTACTTTGCTTCTCTCCACCCATGACCTTTTCCACTGTGGCCAT 1249
 8R←—————
 CAGGACTTTCCCTTGACAGCTGTGTACTCTTAGGCTAAGAGATGTGACTACAGCCTGCCC 1309
 →9F
 CTGACTGTGTGTGCCAGGGCTGATGCTGTACAGGTACAGGCTGGAGATTTTCACATAGG 1369

 TTAGATTCTCATTACGGGACTAGTTAGCTTTAAGCACCCCTAGAGGACTAGGGTAATCTG 1429

 ACTTCTCCTTCCTAAGTTCCCTTCTATATCCTCAAGGTAGAAATGTCTATGTTTCTAC 1489

 TCCAATTCATAAATCTATTACATAAGTCTTTGGGTACAAGTTTACATGATAAAAAGAAATGT 1549

 GATTGTCTTCCCTTCTTTGCACCTTTGABATAAAGTATTATCTCCTGTCTACAGTTTA 1609
 9R←—————
 ATAAATAGCATCTAGTACACATTCA 1634

 3'UTR TTTTGTGTTGGATACTGTGTTAGGTGCTGGAGGAAAAAGATGAATAGAACATC 1688
 TTCTATGTACTTGATGCGCTCACAGTCTGGTTGTAGAGACTGTCACATAAACATTTTCATC 1748
 CCAATTCATTTATTGTTTCATTCTTCAGCCAATATATATTGAGTTCTTACTCTGTGCCA 1808
 AGAACTGTACTACATTTCTGGGATTAAGTGGATATAAGGAGATCTCAGTGTTAATCTGC 1868
 CTGAGGGGAGACTAAATTAAAGTGACATGGAACTTGGGTCTTGAAAAACATTTTAAGGTT 1928
 ATTTTTCTTTTCTCTCTCTCTGCTCTGTCTTTCTCTCTTTCTGCTCAGGGTCTCCCTC 1988
 TGTGCCCAGGCTGGAGTCAGTGGCACTCATAGCTCACTGCAGCCTTGATCTCCTGGGT 2048
 CAAGAGTTCTTCCACCTCAGTCTCCTAAGTAGCTTGGACTACGG 2108

【図12】

カテプシン-K配列：エキソン-イントロン境界

(A) 5' 非翻訳配列

[SEQ ID NO.2]

```
GCTTTGGCTC CCAAAGGCCT GGGATTACAG GCGTGAACCA CTGCGCCTAG
CCTGTTAGCA GCTCTTAAAA TCCAGAGGCA TAAGCCTGTA TTTTGTAGGG
TTTATGCATG GAATCCAGCT AGAAACTGAG TCTATTACAG ATCCCATTTA
TTATCCTTTC TATTCCAAGA AGCCTTTTTT TCTCCTTCCC CACATCTGTT
TATGGAAGAA AATGAAGTTT GGGGTGTGGT TTGAGGAATC AGCTAGATTC
TTATGATCTG TCACATGCTT GGATGTTGGG GAAGCATTTG GAGAAGCTCA
TGTGACTTGT CCTAGATTGG GGATTTTAAT TGAGACAGAT GATGTTTATC
GGGCATCCCA CCACCTGAGA GTTTTAGCAA CAGAGTCACA TGTGAGTCCA
TCAGAACTTA CGGCATTGAT TCAAGTGCTG TCATAAATAA CCAGGACTGC
TGTTTTTGGT TACTTTTAAA GACAGTTTCA TCTGGACTTT CTGGGCATAT
CCTCCTTCAG CAAAACCACA TTAGGCTGGG AAAACTATTC TGCCTGGAAG
TAATGACAAC TTGCAACCAA CAAGCTTATA AAAATACAAA GAATTCTGGA
GCCTATGGCT TCCATTACAT TATTCTTTTA TAGCCTTTTA TGTTCAATTAC
CGCATCCCAG AGGTGAGAGT CAGACACAAA TATGAAAATA GGTTCATG
TTGGAGAGGT AAATCCTAAC AGGAAAGGGG TAGGAAAAGA TATAATCCCC
CAATATTAAA ATAAAGATAT TGAAGAAGAA GGATGGGAGA GACTAGGGCT
GTGTCCTTCC TTTTACTCAC CAAAAGAGAA AGTAAGCTCC TATTTGAGTC
AATAGATATT GAGGTCTTGT TATTGCCCAC CAAAGACAGT CTTGTGAGAC
TAAATAGCTA GTAATCCCTT ACCCTGGCAC ACATGCTGCA TACACACAGA
AACACTGCAA ATCCACTGCC TCCTTCCCTC CTCCCTACCC TTCTTCTCT
CAGCATTTCT ATCCCCGCCT CCTCCTCTTA CCCAAATTTT CCAGCCGATC
ACTGGAGCTG ACTTCCGCAA TCCCGATGGA ATAAATCTAG CACCCCTGAT
GGTGTGCC
```

(B) エキソン1 [SEQ ID NO. 3]

CACACTTTGCTGCCGAAACGAAGCCAGACAACAGATTTCCATCAGCAG

(C) イントロン1[SEQ ID NO. 4]

gtaacgtttg caacttcta gatcttttag cttttcattc ctgtcaattc
 tctgagtatt agggatgtag tgacttgagg atcacaataa acttttagcc
 tctgcagatg aaaacagaga tgcacttctt aggtcattcc ctggctaaat
 aaaatctgcc tggaaatctg tagaattcct tgtatgattt atatataac
 atacatgatt gttagtataa gcaaagtata tagggaatca tttcccatc
 cttcaagagt ggcctttctg cagtgttttc tactttggcc aacaaggatc
 aaaacgggta actccttagt gaggaggagg agagtgggtat ggggaggtag
 tagctcagtg cttcctgttc actgagacat ctcaaagccc ttaacactct
 agttttttaa tgtcctactg gacattttgc cagtttgcaa aattacatgt
 aatgggacta taagcaattg tgtaagccat atgtcatgct gcaggctgca
 aattgtttct aaaatggagg atttgtaatt aagaaagcca atgcaagaaa
 tgagtgaagc taactagagt aaacttatga aaagctgtga atttcatcat
 catagaacat tgcttttcag tctgaacatt cttctaaca accttggatc
 tgaggcttct tgtcctttgc ggcagccaca gtgggttttt gttgttaggg
 gaaaataaaa aaccttgccc gcagcatctg gttaagatta gggcagtttc
 ctgcctaagg agggaaggga gagaaaaagg aagaagaaat gcataaggag
 aatgaggaga tatacaatgt ctcagaaaac aggaaacatt gtcctatttt
 cccttgtcct cttctgacaa gatctgggaa agtaccagaa tttaggcacg
 aaagagaaga acgcctcgaa gaaatgatca ggaagcaaaa cttagacgga
 aatctctcct ttgtgtattc tgaacccac taccaccttg ctatttgtct
 gtctccaagc ctgctagggg ccctggagga aacgcactga gccattctg
 attgtccagt ttctatcccc catttctggg tgtgtacgtg tgtgtgtgtg
 tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgagagagag agagacagag
 agagaaacag agagagtgtg tgttgcctaa atctcccgag agagagagag
 agagagagag agagagagag agagagaaaa gagagaaatg gctaaatccc
 cctagatcaa agtccttgga accagatgta ccagcatcct atctaaacac
 aggccctcc tgactatcat tgttttatca ccctttttcc gtctaccttt
 ctcttcctca taaagcctag ttttcctctg tttccctgcc aaatggaaga
 gttttcccta actacattct tctgcag

(D) エキソン2 [SEQ ID NO. 5]

GATGTGGGGGCTCAAGGTTCTGCTGCTACCTGTGGTGAGCTTTGCTCTGTA
CCCTGAGGAGATACTGGACACCCACTGGGAGCTATGGAAGAAGACCCACA
GGAAGCAATATAACAACAAG

(E) イントロン2 [SEQ ID NO. 6]

*gtgcctgggg tcctggaggg ggcattggcag gaaggctgag acctgagctc
tctcatctta gcttcagac tcccttcttc aatccaaatg ctttattcca
agcaaatcag tccctcttcc ctaactcatg ttaacatacg gttttcattc
ctatgcttca atcatctct tgtcaaactt gtattccttc ccttttggtt
tataagtgtg taacattcct cttttgggaa gagtcccaag attaatgctg
ttaatccata agcaattttt ctgtctctcc agagcttgtg tggttgttta
catattatct ctcttcttgc aggctcttaa ttccatggtt agttcccga
ctaaactgta aacttttatg attgtgagtt tcctttattc tcctaaaacc
cttcacaafa ttacatatga actgtagaca gtctatacaa gtactgacta
tgctttgttt ag*

(F) エキソン3 [SEQ ID NO. 7]

GTGGATGAAATCTCTCGGCGTTTAATTTGGGAAAAAACCTGAAGTATATTTCCAT
CCATAACCTTGAGGCTTCTCTTGGTGTCATACATATGAACTGGCTATGAACCACC
TGGGGGACATG

(G) イントロン3 [SEQ ID NO. 8]

*gcaagtatag cttcagctcc tgtcccacct gcaccatttg ctttagttcc
ctgctgatgc ctggcctctt tcttctttgt ctttag*

(H) エキソン4 [SEQ ID NO. 9]

ACCAGTGAAGAGGTGGTTCAGAAGATGACTGGACTCAAAGTACCCCTGTCTCATTC
CCGCAGTAATGACACCCCTTTATATCCCAGAATGGGAAGGTAGAGCCCCAGACTCTG
TCGACTATCGAAAGAAAGGATATGTTACTCCTGTCAAAAATCAG

(I) イントロン4 [SEQ ID NO. 10]

gtactctcct ttcttctggg tgtgcatatg taatctggca tgaccttttc
ctttttctgc tgctttgttc ttgaggtgaa agggcaccag gaaaagaggg
caaggaatta aggtacatct cccattccc attctgttat ttaacctcat
ttgtttctgt acatttgggt tgtttctgggt ttttcttttt cttttccctt
tttttttttt tttttttttt gagatagagt ctactctgt cgcccaggat
ggagtgcagt ggtgcaatct tggctcactg caacctacac ctcccggtt
caagcgattc tctgcctca gcctcctgag tagctgagat tacaggcagc
cgccactacg cctggctaata ttttctatct ttatagagat gcgtttttac
catgttgggc aggtctgtct tgaactgacc tcaggtgatc cacctgcctc
agcctcccaa agtgctggga ttagagtcac gagccatcgc ggcctggttt
ttctttatta caaatagtgt tgcaataagc acccttgtgc atatgttttt
gtgcacatgt acaaatatct atgcaaaaata agtcctaaaa ttggaattgt
taggtcaciaa ataatccttt ccccccccc aaattttttt tttttttttg
agacagcgtc tctgtcaccg aggtctgggt ccagtggcgc aatcatgggt
cactgcagcc ttaacgtctc aggtcgaagt gattctccaa cctcagcctc
cctagtagct gggaattaga agcacatgcc accacaccca gotaattttta
aaaaattttt tgtagagac aggtttttgc catgctaccc aagctggtct
caaattcctg ggctcaagca atctgcccgc ttcggcctcc caaagtgcta
ggattacaga catgagccac catgcccagc ccaaaaaagt ttttgcaatc
ttacattctt actagcatga gaatgtcagt tttttcacia cccaacaac
acaggattgt atcagcaaga taaacaattg atttaacgtt catttaacaa
acactttttg acccccagaa cctaccagat gcagtgttag gcagcagaga

図15の77"を

ctcaagatga ctaagacaca acctgtgtcc tcaggaaatc tcaatctaaa
 aaaatagaac aggaaagaaa gaaaaatcta caatctagct gcacaaacaa
 taatagctaa tactttttga gatatttttg tttgtcagga acttcttaac
 tctttacatg agttttaata tttaatccct tataacaata ttttatgcat
 agagaaactg agacacaggg aaatttagta acttaccggg ggtcacatag
 ctactgggtg gcaaagtcag ggtagctcc caggacaaat gcctccacag
 ctggtactgt gctctgcttt actgtagcta atagtaaaaa tggtagcaaa
 aatcaatagc agtagaacag tgcaacagat attaagcgga agaggaagac
 tcacaacaat gacaacattt gtgctgaaat ttttaagaac acatggaatt
 tccttcagcc gggtagagag aagatataga aatgtaaaca ccaaagattc
 atagttttctc tgtatccctt tcag

(J) エキソン5 [SEQ ID NO. 11]

GGTCAGTGTGGTTCCTGTTGGGCTTTTAGCTCTGTGGGTGCCCTGGAGGGCCAA
 CTCAAGAAGAAAACCTGGCAAACCTCTTAAATCTGAGTCCCCAGAACCTAGTGGATTG
 TGTGCTGAGAAATGATGGCTGTGGAGGGGCTACATGACCAATGCCTTCCAATATG
 TGCAGAAGAACCGGGGTATTGACTCTGAAGATGCCTACCCATATGTGGGACAG

(K) イントロン5 [SEQ ID NO. 12]

gtgagattgc tccacacaat tatacagctc tgttggctcc tccttcccca
 gcatgatgtt ttgtactgga aacaattcca gaaatactgt tttctgttat
 cctatcctgc tttcttgatg gaataatttc ccacagaagg ccaagaagat
 ttccacaatc tgggggaatt tagggagctt aagctactat agctcctatt
 tgcattctctg ccatggagag aaaacagagg ctaggctacc taccccatag
 acttccgagc tgggttctat aaccctctgc tcaattcctc actccacaa
 caaaccacaa aaccaccat gctattttca caaattgtgt ggctttattt
 tatatgatct cagtgtgagt tttcagaaca tttcagcaaa ttatgtaagt
 ttacatgcta acatctataa aatgagagaa aaaacaagtt gcttcatata

図 16 の つ づ き

agagataagg gattaactca gttcctcctg catgatcctc tagtcatagg
aaggaaatca tatctgaaag ggaggcaacc tgagggggtt tttatacaca
tagggctggg tctgatagac aatataatgt agggccttca caacagaaac
ctctgaaaca gggacagcaa gtttgagaat aaaaatgatg gctactgtgt
tctaagccgt gtccttagtg ctttttttct ttttcttttt ttcatttaat
ctcataacaa ctctgttagg tagacttata ttgaatgtat aggtgaggaa
atggacactt aaggagataa gacagtataa ttcataccac tagtatgtaa
caatgtaaga tgtatctacc agggatgttt atcttctgca aacattccta
ggtatatctc ccatgcacat gtgcaagaat ttcttactag gatataatgc
cttggaactg aattgtctgg gtcttagggg atgtctgtct tcaactttac
tacacaatgt caaattgttt gccaaaatat ttggaaaaat ttatacctgc
aatgtgtaag aaatcccctt caatcacctt tttatcagta tgtttatctg
gccatttgca tttcttcttc agtgaattaa ctgtttttat ctcttgctca
tttggttttc tttttatttt ttgaaatag ggtcttactc tgttgcccaa
ggctggagtg tggtgataga gtcataagctc actgcagcct ccacttcagg
gctcaagcaa tcctctcgcc tcagcctccc aaatagctag gatataagtg
catgccatga tgcccaccaa tttcaaaaaa cctttgaaat tttttttttg
taaaagctag gcatgggtggc tcatgcctgt aatcccagca ctttgggagg
ctgaggtggg aggatcgctt gagcccagga atbaggagtc ggcctgatag
aacatagcaa gacctcatct ctacagaaaa aattttttaa agtagccagg
tatgatggcg tgcatagttc tagctactcc ggaagctggg tgggaggaca
acttgagcct gggagttaa ggctgctgtg aactgtgate atgtcactgc
tctctagcct gggtgacaga gtgagaccct gtccccaaaa acaacaaccg
tttttttttg tagagacatt gtctcgctat gttgccaagg ctagtctcaa
actcctgggc tcaagcaatc ctcccacctc ccaaagtgtt gggattatag
atgtaagcca ccatgcctgg cctacccttt tttttttttt ttgaaatgga
agttttgctt ttgtcaccta ggcttgagtg cagtggcgcg atcttggctc
actgcaacct ccacctcctg gattcaagca attctcctgc ctgagcctcc
tgaatagctg ggattatagg caccgcgaac cagccccggc tagtttttgt
atttttagta cagacagggg ttcacatgt tggccagctg gtcttgaacc
cctgacctca ggtgggtccg ccgcctcggc ctcccaaat gctgggatta

17 の つ 7 " 2

aaagtgtgag ccaccatgcc ccaccctta ctcattttta attggattgt
 tttttctctt tcttagcgat tcttaaaagt ttaaagagaa tatttggata
 caatactatg tatttaaaag ttgaggtctg tctttccatt cttctcacga
 tgtctttcaa tctagaaaag ttaattttta taggcctggc gcggtggctc
 acgcttgtaa tcccagcact ttgggaggct gagatgggtg gatcacaagg
 tcaggagatg aagaccatcc tggctaaccat gggtgaaacc ctgtttctac
 taaaaatata aaaaaattag ctgggcgtgg tggcagggtgc ctgtagttcc
 agctactcgg gaggctgagg caggagaatg gcgtgaaccc gggagggtgga
 gcttgcatg agccgagatt gcaccactgc actccagcct gggcaactga
 gcaagactgc gtttcaaaaa aaaaaaaagt taattttaat atagtataat
 tagtaaaagg attaattttc cctttgcaat ttttgtaatg tgttttattc
 gtttatgaat ggagaaagggt aagaaaaaat aaaattttaa aaagaagaga
 tgtggccagg tacggtggct cacacctata atcccagtag tttgggaggc
 tgaggcaggc agatcacttg aggtcaggag tttgagacca gctgggataa
 catggtgaaa ccccatctct actaaaaata caaaaattag ccagggtgtga
 ttgctgcacgc ttgtaatccc agcaggctga ggcaggagaa ttgctcgaaac
 tcaggaggca gaggttgcag tgagccaaga tcatgccatt gcactccagc
 ctgggtaaca gagactctgt ttcaaaaaa taaaaagata aaaaaggag
 agatctgata gggcgccag ataaacattt taaaggggat ggtattataa
 gtttgttccc agcataatgc caggttattc tgacttttaa gtatcatcac
 ataatatctt tttgagtcaa tttccaagat attctgtttc acttgtaatt
 ctgtgtaatt tttggcacca ggaggcatca gggatttgga gcacatggca
 gaaacaaagg catcttgaaa aatatcaagg cagtagacca ctgtaatctt
 aaaatggcat atcaaagtgt gctattgctg ttaatattta gataatgtta
 gataatgtat ttttttagag ggtatctcac tatcttgca aggctggagt
 agagtggcta ttcacagcat gatcacagta cactaaaggc tcaaactcct
 gggcacaaac aatcctcctg cctcagcctg ctgagtagta gataataagt
 tcttgtggat gcaaccttag ggttctgaag gggtagtctg taggaaaatg
 aattgtgtaa aagaatacac caccttaaca tgggctatta ttcgattcca
 taattgtggc ttgccaatga aacattgcta actacctgta aaatatagt
 ttggaagtca taggctaaat tgctaagttc tttaatctat tttagtgtct

図18の77"キ

tgttatgtac ttttatattt tgtctttgat gagagcaciaa ggatcacacc
agttcccttg atataggtgc agagggccca ggtcttccct ctagctaagc
cttgcccttg gcctcctacc cacacagcag ctggtgcctt cctgccccct
gaggctaata catactatgt ggccagaaga tggtttatgc tttttaaaaa
aatcttattt cagaaatctt tccctactgt tttcctccca catttatgtc
ttaaaacacc tgtaggggat tttttttttt tttttttttt tgagatggag
tctcgctctc gccaggtgc gagtgcattg gcgcgatctt ggctcactgc
aaggtctgcc tcccaggttc acgccattct cctgcctcag cctccccagt
agctgggact acagggcgcc gctaccacgc ctggctaatt tttttgcatt
tttagtagag acaggggttc actgttagcc aggatgggtc cgatctcctg
acctcgtgat ccaccctcct cagcctccaa agtgctggga ttaacaggca
tgagagccca ccgcactggc ctgtatttgt gaggaagaac agaccctctt
tagaagccct agactgctgc ctctgttagt tcaactggcat cactcaaaat
attgggttag tttcttactc actgagttgg tttttatgtg tgggtggaagg
cggaatcct cttttcatac tcgttctcat tgcctattgc tttgtcctag
tctattaca atcttgtttc ttccag

(L) エキソン6 [SEQ ID NO. 13]

GAAGAGAGTTGTATGTACAACCCAACAGGCAAGGCAGCTAAATGCAGAGGGTACAG
AGAGATCCCCGAGGGGAATGAGAAAGCCCTGAAGAGGGCAGTGGCCCGAGTGGGAC
CTGTCTCTGTGGCCATTGATGCAAGCCTGACCTCCTTCCAGTTTTACAGCAAAG

(M) イントロン6 [SEQ ID NO. 14]

gtaagaagct gctgatccta tacagcactg tcttttatga taaaaacttg
atggtttctc gaaggacctt gggatatttc agtacttagt ttttgtattc
acatggaggt ggccagagag aaattaacaa ctgctgcagt atggagcagc
atctctgtgg taaacctcc tgacacggat ggaattcttc aaacagtctc
ctagactggg agatcccaca gggtagacct tggattgcat agagcctcac
gctggtagtt tgtattctag

(N) エキソン7 [SEQ ID NO. 15]

GTGTGTATTATGATGAAAGCTGCAATAGCGATAATCTGAACCATGCGGTTTTGGCA
GTGGGATATGGAATCCAGAAGGAAACAAGCACTGGATAATTAAAAACAG

(O) イントロン7 [SEQ ID NO. 16]

gtaatgatgg gaacactact tttgttattc agtcaccctt ttaacactca
acctcacctc cagcttcccg atattccttt ctctgtccca aatcaagaaa
aaattatct cagagtcttc acttctatct tctcagtcag aggctcttaa
ttctcagctc gacacttaat ggccagtggt ttagtccatt ttgcattgcc
acaaaagaat acccgagact gggtagttta taaagaaacg aggtttgttt
ggctatacaa agcgtggcac tagtatctgc tcagcctctg atgaggcctc
agagctttta ctcatggcag aaggcaaaag agggagcagg catgtcacat
agtgagagag ggagcaagag agagagggag gtgccgactc tttaaagaac
cagctcttgc atgaactaat agagtgagaa ctactcatc accaaggcga
tggcaccapg ccattccatg aggaatccac ttcataacc caaacacctc
ccactatgcc ccacctccca cattggggat cacatttcag catgagactg
ggaggggaca cacatccaaa ccatatccgc cagacaatag tgctcaatta
tgtgctgggc agatgctccc tgtgtgcaag gtgcttagtg acatacataa
accaacgagc agatgacacc ttcagtgagc tcagagccca ataagacaga
cctaactaac catgagataa agcagtacaa agaaccagca ggagctttgg
aattacgtat ttttactttc ttttgtctct aatgtgatca gtttcttaga
tggtttccat tagcaatctg tctttaacag taggggagca gctttaaggg
tttaatatct cttttgaaca gtttttttcc ttcaaaatac acttaagata
cacgtatata agaacttgcc aaagattgtg aagagaaaca ttttttagaa
ataagatata aacaaaaaaa gttagtgtta ctttcctatg ttggggaaca
aagaaaactc cagggtagct tgcctcccat ttctcttttag caccttgtga
cttttgggga ggggcagatt gataacaatt atagttttcc tttcctggct
gatcaccatt aacctggtag cagcactggc taaatctcct gtccttagtg
ccctccaagg agcaggagcc ctagactctg ggtcgctgac agactcacgc

図20のつぎ

agtgggtgttg ttcaaacctg aagcaacttt ttatatcaca gttccaactc
 aagggtgaacc tgagcatctt cccaagtctc ccacagcttc tgtcctgtgt
 tgtcccttct cttgactccc aggtccaagc acttaccctg ttctttcatg
 atcaggtacc atgtgtggag atagcttcca agagagctgg gaggaagaaa
 ggacacaccc gggcaggatc aggaacactg ggggcccctg gagaagggga
 gagtggggga ggggtacaggt tttaaataaa atgtgttggt aattagagaa
 ttgctggttg gggaaagagg tctgaaaaca attcaggaag ataaacaaga
 caatctctcc tctctcctct tcttcacgtc gtctctcttg tcttctagtc
 tctgtactca tttccttagt aatctcatcc actctcatag tttcatccat
 ctctcctatg gggtttacc ccaaatacag atcaccagct tcagcctcct
 tcttatgtc taaactcaca ttttcaagat taatattccc caaatacagc
 tctgatcata tcaactctcc actcaaaatc cctcactggc tcctcacgat
 gatgggtcac agagtaaagg tgaagctttt taaccttgca gtaaaggtaa
 ttcaacctga tctcaatctg cctttccaga catctctccc actacacct
 gttaggcaca ctgcttttca gctacatgat cctaacagtg cccacactt
 tcctgcctct gttgttcatt tcacaccctt ccactggcat ccccttccc
 caggtcgaaa ttctacttag ccttttggt cagctcaa at gccacctctt
 acatcaagcc tctaagattc tcttgatcag aaggaatctt tccctccttt
 gatacctaca gtattatgcc ttctccctat ttcttgactt taaactcttt
 aaagttaaaa aacatcatat tcatttttgt gtaccatcag tacctcgac
 aatactcagt aaatatatta atgaataaat aaactgagag tactaagtat
 ttttcttgat tggctttaca g

(P) エキソン8 [SEQ ID NO. 17]

CTGGGGAGAAACTGGGCAAAACAAAGGATATATCCTCATGGCTCGAAATAAGAACA
 ACGCCTGTGGCATTGCCAACCTGGCCAGCTTCCCCAAGATG End Coding

【図22】

(Q) 3' 非翻訳配列 cDNA

[SEQ ID NO. 18]

TGACTCCAGCCAGCCCAAATCCATCCTGCTCTTCCATTTCCCTCCACGATGGTG
CAGTGTAACGATGCACTTTGGAAGGGAGTTGGTGTGCTATTTTTGAAGCAGATGTG
GTGATACTGAGATTGTCTGTTCAGTTTCCCATTTGTTTGTGCTTCAAATGATCCT
TCCTACTTTGCTTCTCTCACCCATGACCTTTTTTCCACTGTGGCCATCAGGACTTT
CCCCTGACAGCTGTGTACTCTTAGGCTAAGAGATGTGACTACAGCCTGCCCTGAC
TGTGTTGTCCCAGGGCTGATGCTGTACAGGTACAGGCTGGAGATTTTACATAGGT
TAGATTCTCATTACGGGACTAGTTAGCTTTAAGCACCCCTAGAGGACTAGGGTAAT
CTGACTTCTCACTTCCTAAGTTCCTTCTATATCCTCAAGGTAGAAATGTCTATGT
TTTCTACTCCAATTCATAAATCTATTCATAAGTCTTTGGTACAAGTTTACATGATA
AAAAGAAATGTGATTTGTCTTCCCTTCTTTGCACTTTTGAAATAAAGTATTTATCT
CCTGTCTACAGTTTAAATAAATAGCATCTAGTACACATTCA

(R) cDNAの範囲外の3' 非翻訳配列

[SEQ ID NO. 19]

TTTGTGTG GATACTGTGT TAGGTGCTGG AGGAAAAAAG ATGAATAGAA
CATCTTCTAT GTACTTCATG CGCTCACAGT CTGGTTGTAG AGACTGTCAC
ATAAACATTT CATCCCAATT CATTTATTTG TTCATTCTCT CAGCCAATAT
ATATTGAGTT CTTACTCTGT GCCAAGAAGT GTACTACATT TCTGGGATTA
AGTGGATATA AGGAGATCTC AGTGTTTAAT CTGCCTGAGG GGAGACTAAA
TTAAGTGACA TGGAAACTTG GGTCTTGAAA AACATTTTAA GGTTATTTTT
TCTTTTCTCT CTCTCTCGCT CTGTCTTTCT CTCTCTTTCT TCAGGGTCTC
CCTCTGTTGC CCAGGCTGGA GTCAGTGGCA CTCATAGCTC ACTGCAGCCT
TGATCTCCTG GGCTCAAGAG TTCTTCCCAC CTCAGTCTCC TAAGTAGCTT
GGACTACGG

-1108 GCTTTGGCTC CC~~AAAGGCCT~~ GGGATTACAG GCGTGAACCA CTCGGCTAG CCTGTTAGCA GCTCTTAAA TCCAGAGGCA -1029
 -1028 TAAGCCTGTA TTTTGTAGGG TTATATGCAATG GAATCAGCT AGAAATGAG TCTATTACAG ATCCCATTTA TTATCCTTTC -949
 -948 TATTCCAAAG AGCCTTTTTC TCTCCTTCCC CACATCTGTT TATGGAAGA AATGAAGTTT GGGCTGTGCT TTGAGGAAATC -869
 AP3 Pu.1
 -868 AGCTAGATTG TTATGATCTG TCACATGCTT GGAATGTTGG GAAGCATTTG GAGAGCTCA TOTCACTTCT CCTAGATTGG -789
 -788 GGAATTTAAT TGAGACAGAT GATGTTTATC GGCATATCCA CCACCTGAGA GTTTTAGCMA CAGAGTCACA TGTGATGCCA -709
 -708 TCAGAACTTA CGGCATTGAT TCAAGTGTCTG TCTAATATTA CCAGGACTGC TGTGTTTGGT TACTTTTAAA GACAGTTTCA -629
 -628 TCTGACTTTC CTGGGCATAT CCTCCTTCAG CAAAGCCACA TTAGGCTGGG AAACTATTC TGCCTGGAAG TAATGACAAC -549
 AP3-RV SPI/H-APV-1
 -548 TTGCACCCAA CAGCTTATA AAAATACAAA GAATTCGGA CCTATGGCT TCCATTACAT TATTCCTTTA TAGCCTTTTA -469
 -468 TGTTCATTAC CGCATCCCAAG AGGTGAGACT CAGACACAAA TATGAAAATA GOTTTCAAATG TTGAGAGAGGT AAATCCTAAC -389
 -388 AGGAAAGGGG TAGGAAAAGA TATAATCCC CAATATTTAA ATAAAGATAT TGAAGAAAGA GATGGGAGA GCTAGGGCT -309
 PE13 " PE13
 -308 GTCTCCTTCC TTTTACTCAC CAAAGAGAA AGTAAGTCC TATTGAGTC ATAGATATT GAGGTCTTGT TATTGCCAC -228
 PE13-RV AP1-RV AP1
 -227 CAAAGACAGT CTTGTGAGAC TAAATAGCTA GTAAATCCCT ACCGTGGCAG ACATGCTGCA TACACACAGA AACCTGCA -149

(12) 23 の 77' まで)

-148 ATCCACTGCC TCCTTCCTC CTCCTACCC TTCCTCTCT CAGCATTCT ATCCCGCT CTCCTCTTA CCAAAATTT -69
 PEA3-RV SPI-RV

-68 CCAGCCGATC ACTGAGCTG ACTTCGCAA TCCGATGGA ATAACTAG-CACCCCTGAT GGTGTGCCA CACTTCTG +12
 H-APP-1 Ets-1RV ATに富むセチーフ エキソン1→

表 I I
 ヒトカテプシンK遺伝子のエキソン-イントロン結合

No.	CDNA (bp)	ドナー	アクセプター	イントロン 大きさ (bp)	切断された アミノ酸
1	48	GCAGGtaacgtttgcaact....ctacattcttctcagGATG		1427	非コードディング
2	169	CAGGtgccctgggtccctg....ctatgctttgttttagGTGG		462	Lys40/Val41
3	292	CATGgcaagtataagcttca....tcttcttctgtcttagACCA		85	Met81/Thr82
4	448	TCAGgtactctctcttctt....ctgtatcccttctcagGCTC		1624	Gln133/Gly134
5	667	ACAGgtgagtgaatttget....gtttcttccagccagGAAG		4326	Gln206/Glu207
6	833	AAAGgtagaagctgctga....tagtttcttcttagGTGT		270	Gly262
7	939	ACAGgtaatgatgggaaca....tgattggtcttctacagCTGG		2270	Ser297

エキソンおよびイントロン配列は、各々、大文字、小文字で示される。

[26]

	1		50
HumcatKHW GLKVL LLPV SFA.LY PEEI	LOTIHWELWKK	THHKQYNNKV
RabOC-2HW GLKVL LLPV SFA.LH PEEI	LOTQWELWKK	TYSKQYNSKV
HumcatSHKK LVCVLLVCSS	AVAQLHKDPT	LDHWHHLWKK
HumcatLHNFTL ILAAPCLGIA	S.ATLTFDHS	LEAQWTKWKA
HumcatH	HWATLP L LCA	CAWLLGV PVC	CAALSVNSL
HumcatUHWQLWAS	LCCLIVLAAA
HumcatDMQP SSLLPLALCL	LAAPASALVR	IPLIKFTSIR
HumcatEMKT LLLLLLVLE	LGEAQGSLHR	VPLRRHPSLK
HumcatGMQP LLLLLAFLLP	TGAENGEI..ICGRK
	51		100
HumcatK	DEISRRL.IW	EKNLKYISIH	NLEASLGVHT
RabOC-2	DEISRRL.IW	EKNLKHISIH	NLEASLGVHT
HumcatS	EEAVRRL.IW	EKNLKFVMLH	NLEHSMGHHS
HumcatL	EEGWRRR.VW	EKNMKHIELH	NQYREGKHS
HumcatH	EEYHHRLOTF	ASNWRKINAH	N....NGNHT
HumcatB	RSRPSFHPVS	DELVNYVNKR	NTTWQAGHNF
HumcatD	EDLIAGPVS	KYSQAVFAVT	EGPIPEVLKN
HumcatE	SEFWKSHNLD	MIQFTESCSH	DQSAKEPLIN
HumcatG	SRPHSRPYMA	YLQIQSPAGQ	SRCC.....C
	101		150
HumcatK	TGLKVPLSHS	RSNDTLYIPE	WEGRAP.DSV
RabOC-2	TGLKVPPSR	RSNDTLYIPD	WEGRTP.DSI
HumcatS	SSLRVP.SQW	QRNIT.YKSN	PNRILP.DSV
HumcatL	NGFQ...NRK	PRKGVFQEP	LFYEAP.RSV
HumcatH	L.WSEPQNC	ATKSNYLRGT	..GPYP.PSV
HumcatBGGPK	PPQRVHFTD	LKLPASFDAR
HumcatD	TVVFDTGSSN	LWVPSIHCKL	LDIACWIIHK
HumcatE	TVVFDTGSSN	LWVPSVYCT.	..SPACKTHSR
HumcatG	NVTLC.....AHNIQRR	ENTQQH..IT
	151		200
HumcatK	CHAFSSVGAL	EGQLKKKTGK	LLN..LSPQN
RabOC-2	CHAFSSVGAL	EGQLKKKTGK	LLN..LSPQN
HumcatS	CHAFSAVGAL	EAQLKLKTGK	LVS..LSAQN
HumcatL	CHAFSATOAL	EGQMFRTGR	LIS..LSEQN
HumcatH	CHTFSTTGAL	ESAIAIATGK	MLS..LAEQQ
HumcatB	CHAFGAVEAI	SORICIHNA	HVSVEVSAED
HumcatD	HYGSGSLSGY	LSQDTVSVP	QSASSASALG
HumcatE	QYGTGSLSGI	IGADQVSV..E
HumcatG	QYNQRTIQND	IMLLQLSRR.VRRNRNVNP
	201		250
HumcatK	MTNAFQYVQK	NRGIDSEDAYPYVQGEE
RabOC-2	MTNAFQYVQR	NRGIDSEDAYPYVQDE
HumcatS	MTTAFQYIID	NRGIDSDASYPYKAMD
HumcatL	HOYAFQYVQD	NGCLDSEESYPYEATEE
HumcatH	PSQAFEYILY	NRGIMCEDTYPYQCKDC
HumcatB	PABAHNF.WT	RKGLVSGOLY	ESHVGCPRYS
HumcatD	IAAKFDGIL.	..CHAYPRIS	VNNVLPVFDN
HumcatE	VDAEFDGIL.	..GLGYPSLA	VGGVTPVFDN
HumcatG	RPGLCTVA.	..G..WGRVS	HRRGDTLRE

(図 26 の 77 " 2)

	251				100
HumcatK	SCH.....	.YNPTCKAAK	CRGYREIPEG	N.EKALKRAV	ARVGPVSVAI
RabOC-2	SCH.....	.YNPTCKAAK	CRGYREIPEG	N.EKALKRAV	ARVGPVSVAI
HumcatS	KCQ.....	.YDSKYRAAT	CSKYTELPG	R.EDVLKEAV	ANKGPVSVGV
HumcatL	SCK.....	.YNPKYSVAN	DTGFVDIPK.	Q.EKALMKAV	ATVGPVSVAI
HumcatH	YCK.....	.FQPGKAIGF	VKDVANITII	D.EEAKVEAV	ALYHVPVSAF
HumcatB	TPKCSKICEP	GYSPTYKQDK	HYGYNSYSVS	NSEKDIMAET	YKIKHPVKGAF
HumcatD	DAQPGGELHL	GGTDSKYKYG	SLSYLVNTRK	AYHQVHLOQV	EVASHITLCK
HumcatE	EGGAGSELIF	GGYDHSKFSG	SLNWPVETKQ	AYWQIALDNI	QVGTVMVPCS
HumcatG	RRQ.....ICVGDOR	RERKAAPK..	GDSHGPVLAH

	301				150
HumcatK	DASLTSFQFY	SKGVYYDESC	..NSDNLNHA	VLAVGYGIQ.	...EKNKHWI
RabOC-2	DASLTSFQFY	SKGVYYDENC	..SSDNVNHA	VLAVGYGIQ.	...EKNKHWI
HumcatS	DARHPSFFLY	RSGVYYEPSC	...TONVNHG	VLVVGYGDL.	...HCKEYWI.
HumcatL	DAGHESFLFY	KEGIYFEPDC	..SSEDHMHG	VLVVGYGFEF	TESDNHKKYWI.
HumcatH	EVTQD.FHMY	RTGIYSSTSC	HKTPOKVNHA	VLAVGYG...	.EKNHIPPYWI
HumcatB	SV.YSDFLLY	KSGVYQHVTC	EMHGG...HA	IRILGWGVE.	...HGTTHYWI.
HumcatD	EGCEA...IV	DTGTLNHVGP	VDEVRELQKA	IGAVPLIQGE	YHIPPCKVST
HumcatE	EGCQA...IV	DTGTLNITGP	SDRIKQLQNA	IGAAP.VDGE	YAVECANLNV
HumcatG	NVAHG...IV	SYGKSSGVPPEVFTRV	SSFLPWIRTT	HR....SFKI.

	351				400
HumcatK	IK.....NS	WGENHGNKGY	ILMARNKNNH	CGIAN..LAS	FPMH.....
RabOC-2	IK.....NS	WGESWGNKGY	ILMARNKNNH	CGIAN..LAS	FPMH.....
HumcatS	VK.....NS	WGHNFGEEGY	IRMARNKGNH	CGIAS..FPS	YPEI.....
HumcatL	VK.....NS	WGEWGHGGY	VKMAKDRNH	CGIAS..AAS	YPTV.....
HumcatH	VK.....NS	WGPQWGHNGY	FLIERCK.NH	CGLAA..CAS	YPIPLV....
HumcatB	VA.....NS	WNTDWDNGGF	FKILRGQ.DH	CGIESEVVAG	IPHTDQYWRK
HumcatD	LPAITLKLGG	KGYKLSPEY	TLKVSQAGKT	LCLSGFHGHD	IPPPSGPIAH
HumcatE	HPDVTFITNG	VPYTLSPYAY	TLLDFVDCMQ	FCSSGFQGLD	IPHTDQYWRK
HumcatG	LQMETPL..

	401		428
HumcatK
RabOC-2
HumcatS
HumcatL
HumcatH
HumcatB	I.....
HumcatD	LGDVFIGRYY	TVFDRDNRRV	GFPAEARRL
HumcatE	LGDVFIRQFY	SVFDRGNRRV	GLAPAVP.
HumcatG

【図28】

[SEQ ID NO: 20]

```

H W G L K V L L L P V V S F A L Y P E E
I L D T H M E L L W K K T H A K Q Y N N K
V D E I S A R L I M E K W L K Y I S T H
N L E A S L G V K T Y E L A M N E L G D
M T S E S V V Q K K T G L K V P L S H S
R S N D T L Y I P E W S O R A P D S V D
Y R K K O Y V T P V X N Q G Q C O S C W
A F S S V Q A L E G Q L K K K T G K L L
N L S P O N L V D C V S E N D Q C G O O
Y M T H A F Q Y V Q K N R Q I D S E D A
Y P V V G Q E E S C H Y N P T G K A A K
C R G Y R E I P E G H E K A L K E A V A
K V G P V S V A I D A S L T S F Q V Y S
K G V Y Y D E S C N S D H L N H A V L A
V G Y O I Q E O N E R W I I E N S M O S
N W G N K O Y I L K A R R E N H A C O Y
A M L A S F P E K

```

【図29】

増幅された イントロン	オリゴヌクレオチド配列 (5' to 3')	cDNA 位置	
1	CCGAAACGAAGCCAGACAAC (S)	エキソン1	SEQ ID NO.:21
	GCTCACCACAGGTAGCAGCAG (AS)	エキソン2	SEQ ID NO.:22
2	CTCCTGCTACCTGTGGTGAGC (S)	エキソン2	SEQ ID NO.:23
	CCCAAATTAAACGCCGAGAG (AS)	エキソン3	SEQ ID NO.:24
3	CTCTCGGCGTTTAATTGGG (S)	エキソン3	SEQ ID NO.:25
	GGTACTTTGAGTCAGTCATC (AS)	エキソン4	SEQ ID NO.:26
4	CCAGACTCTGTCGACTATCG (S)	エキソン4	SEQ ID NO.:27
	CACATATGGGTAGGCATCTC (AS)	エキソン5	SEQ ID NO.:28
5 & 6	GAAGATGCCTACCCATATGTG (S)	エキソン5	SEQ ID NO.:29
	GTTACATTATCGCTATTGCAC (AS)	エキソン7	SEQ ID NO.:30
7	GCAAAGGTGTGTATTGATG (S)	エキソン7	SEQ ID NO.:31
	GCCGTTGTTCTTATTTGAGC (AS)	エキソン8	SEQ ID NO.:32

フロントページの続き

(51) Int. Cl. ⁶

識別記号

F I

C 1 2 N 9/64

A 6 1 K 37/54

C 1 2 Q 1/68

C 1 2 N 5/00

B

// (C 1 2 N 5/10

C 1 2 R 1:91)

(C 1 2 N 9/64

C 1 2 R 1:91)

(72)発明者 ジュディス・エイ・ブレイク
アメリカ合衆国20878メリーランド州ロー
レル、マラド・ドライブ9933番

(72)発明者 クリスティーヌ・エム・デブーク
アメリカ合衆国19087ペンシルベニア州ウ
ェイン、ビュフ・ロード667番

(72)発明者 フレッド・エイチ・ドレイク
アメリカ合衆国19343ペンシルベニア州グ
レンムーア、ウォルナット・バンク・ロー
ド24番

(72)発明者 リサ・エム・フィッツジェラルド
アメリカ合衆国20876メリーランド州ジャ
ーマンタウン、オブザベーション・コート
13番 アpartment・ナンバー201

(72)発明者 クレア・エム・フレイザー
アメリカ合衆国20854メリーランド州ノー
ス・ボトマック、グレン・ミル・ロード
11915番

(72)発明者 マキシム・ゴーウェン
アメリカ合衆国19481ペンシルベニア州バ
レー・フォージュ、リパティール・レイン10
番 ビー・オー・ボックス463

(72)発明者 グレグ・エイ・ヘイスティングス
アメリカ合衆国20874メリーランド州ジャ
ーマンタウン、アンセル・テラス13504番

(72)発明者 イーウェン・エフ・カークネス
アメリカ合衆国20832メリーランド州オー
ルニ、リトル・ビスタ・テラス2519番

(72)発明者 ノーマン・エイチ・リー
アメリカ合衆国21163メリーランド州ウッ
ドストック、ウェザーバーン・ロード
10344番

(72)発明者 ジュリー・ロッド
アメリカ合衆国19050ペンシルベニア州ラ
ンスダウン、ウエスト・ボルティモア・ア
ベニュー80番

CATHEPSIN K GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Application No. 60/019,942 filed June 14, 1996, U.S. Provisional Application No. 60/020,273 filed June 17, 1996 and U.S. Provisional Application No. 60/026,083 filed August 26, 1996.

This invention relates, in part, to newly identified polynucleotides and polypeptides; variants and derivatives of the polynucleotides and polypeptides; processes for making the polynucleotides and the polypeptides, and their variants and derivatives; agonists and antagonists of the polypeptides; and uses of the polynucleotides, polypeptides, variants, derivatives, agonists and antagonists. In particular, in these and in other regards, the invention relates to polynucleotides and polypeptides of human cathepsin K, especially genomic sequences of cathepsin K, and most especially promoter and intronic sequences.

BACKGROUND OF THE INVENTION

Bone resorption involves the simultaneous removal of both the mineral and the organic constituents of the extracellular matrix. This occurs mainly in an acidic phagolysosome-like extracellular compartment covered by the ruffled border of osteoclasts. Barron, et al., *J. Cell Biol.*, 101:2210-22, (1985). Osteoclasts are multinucleate giant cells that play key roles in bone resorption. Attached to the bone surface, osteoclasts produce an acidic microenvironment between osteoclasts and bone matrix. In this acidic microenvironment, bone minerals and organic components are solubilized. Organic components, mainly type-I collagen, are thought to be solubilized by protease digestion. There is evidence that cysteine proteinases may play an important role in the degradation of organic components of bone. Among cysteine proteinases, cathepsins B, L, H, and S can degrade type-I collagen in the acidic condition. Etherington, D.J. *Biochem. J.*, 127, 685-692 (1972). Cathepsin L is the most active of the lysosomal cysteine proteases with regard to its ability to hydrolyze azocasein, elastin, and collagen.

Cathepsins are proteases that function in the normal physiological as well as pathological degradation of connective tissue. Cathepsins play a major role in intracellular protein degradation and turnover, bone remodeling, and prohormone activation. Marx, J.L., *Science*, 235:285-286 (1987). Cathepsin B, H, L and S are ubiquitously expressed lysosomal cysteine proteinases that belong to the papain superfamily. They are found at constitutive levels in many tissues in the human including kidney, liver, lung and spleen. Some pathological roles

of cathepsins include an involvement in glomerulonephritis, arthritis, and cancer metastasis. Sloan, B.F., and Honn, K.V., *Cancer Metastasis Rev.*, 3:249-263 (1984). Greatly elevated levels of cathepsin L and B mRNA and protein are seen in tumor cells. Cathepsin L mRNA is also induced in fibroblasts treated with tumor promoting agents and growth factors. Kane, S.E. and Gottesman, M.M. *Cancer Biology*, 1:127-136 (1990).

The gene expression and cellular content of a non-cysteine protease, cathepsin D, in Alzheimer's disease brain showed evidence for early up-regulation of the endosomal-lysosomal system. Cataldo AM, et al., *Neuron*, 1995, 14 (3), 671-680).

In vitro studies on bone resorption have shown that cathepsins L and B may be involved in the remodelling of this tissue. These lysosomal cysteine proteases digest extracellular matrix proteins such as elastin, laminin, and type I collagen under acidic conditions. Osteoclast cells require this activity to degrade the organic matrix prior to bone regeneration accomplished by osteoblasts. Several natural and synthetic inhibitors of cysteine proteinases have been effective in inhibiting the degradation of this matrix.

The isolation of cathepsins and their role in bone resorption has been the subject of an intensive study. OC-2 has recently been isolated from pure osteoclasts from rabbit bones. The OC-2 was found to encode a possible cysteine proteinase structurally related to cathepsins L and S. Tezuka, K., et al., *J. Biol. Chem.*, 269:1106-1109, (1994).

An inhibitor of cysteine proteinases and collagenase, Z-Phe-Ala-CHN₁ has been studied for its effect on the resorptive activity of isolated osteoclasts and has been found to inhibit resorption pits in dentine. Delalisse, J.M. et al., *Bone*, 8:305-313 (1987). Also, the effect of human recombinant cystatin C, a cysteine proteinase inhibitor, on bone resorption *in vitro* has been evaluated, and has been shown to significantly inhibit bone resorption which has been stimulated by parathyroid hormone. Lerner, U.H. and Grubb Anders, *Journal of Bone and Mineral Research*, 7:433-439, (1989). Further, a cDNA clone encoding the human cysteine protease cathepsin L has been recombinantly manufactured and expressed at high levels in *E. coli* in a T7 expression system. Recombinant human procathepsin L was successfully expressed at high levels and purified as both procathepsin L and active processed cathepsin L forms. Information about the possible function of the propeptide in cathepsin L folding and/or processing and about the necessity for the light chain of the enzyme for protease activity was obtained by expressing and purifying mutant enzymes carrying structural alterations in these regions. Smith, S.M. and Gottesman, M.M., *J. Bio Chem.*, 264:20487-20495, (1989). There has also been reported the expression of a functional human cathepsin S in *Saccharomyces*

整理番号 156518

cerevisiae and the characterization of the recombinant enzyme. Bromme, D. et al., J. Biol. Chem., 268:4832-4838 (1993).

SUMMARY OF THE INVENTION

5 Toward these ends, and others, it is an object of the present invention to provide polypeptides, *inter alia*, that have been identified as novel cathepsin K by homology between the amino acid sequence set out in Figure 5 and known amino acid sequences of other proteins such as rabbit OC-2 and human cathepsin O cDNA. Tezuka, K., et al., J. Biol. Chem., 269:1106-1109, (1994).

10 It is a further object of the invention, moreover, to provide polynucleotides that encode cathepsin K, particularly polynucleotides that encode the polypeptide herein designated cathepsin K.

 In a particularly preferred embodiment of this aspect of the invention the polynucleotide comprises the region encoding human cathepsin K in the sequence set out in
15 Figure 1 [SEQ ID NO: 1] or in the genomic DNA (herein "gDNA") in ATCC deposit No. 98035 (referred to herein as the deposited clone).

 In accordance with this aspect of the invention there are provided isolated nucleic acid molecules encoding human cathepsin K, including mRNAs, cDNAs, genomic DNAs and, in further embodiments of this aspect of the invention, biologically, diagnostically,
20 clinically or therapeutically useful variants, analogs or derivatives thereof, or fragments thereof, including fragments of the variants, analogs and derivatives.

 Among the particularly preferred embodiments of this aspect of the invention are naturally occurring allelic variants of human cathepsin K.

 It also is an object of the invention to provide cathepsin K polypeptides,
25 particularly human cathepsin K polypeptides, that cause or are associated with disease, for example, osteoporosis, Paget's disease, Gaucher's disease, CNS inflammation, Alzheimer's disease, hyperparathyroidism, bone degradation, metastatic tumors, rheumatoid arthritis, osteoarthritis, periodontal disease and degradation of bone implants and bone protheses, particularly dental implants.

30 In accordance with this aspect of the invention there are provided novel polypeptides of human origin referred to herein as cathepsin K as well as biologically, diagnostically or therapeutically useful fragments, variants and derivatives thereof, variants and derivatives of the fragments, and analogs of the foregoing.

Among the particularly preferred embodiments of this aspect of the invention are variants of human cathepsin K encoded by naturally occurring alleles of the human cathepsin K gene.

5 It is another object of the invention to provide a process for producing the aforementioned polypeptides, polypeptide fragments, variants and derivatives, fragments of the variants and derivatives, and analogs of the foregoing.

In a preferred embodiment of this aspect of the invention there are provided methods for producing the aforementioned cathepsin K polypeptides comprising culturing host cells having expressibly incorporated therein an exogenously-derived human cathepsin
10 K-encoding polynucleotide under conditions for expression of human cathepsin K in the host and then recovering the expressed polypeptide.

In accordance with yet another object of the invention there are methods to determine drug responsiveness of individuals having or suspected of having a defect in the cathepsin K gene.

15 In accordance with yet another object the invention there are provided products, compositions, processes and methods that utilize the aforementioned polypeptides and polynucleotides for research, biological, clinical and therapeutic purposes, *inter alia*.

In accordance with certain preferred embodiments of this aspect of the invention, there are provided products, compositions and methods, *inter alia*, for, among other things:
20 assessing cathepsin K expression in cells by determining cathepsin K polypeptides of cathepsin K-encoding mRNA or hnRNA *in vitro*, *ex vivo* or *in vivo* by exposing cells to cathepsin K polypeptides, polynucleotides or antibodies as disclosed herein; assaying genetic variation and aberrations, such as defects, in cathepsin K polynucleotides, genes and gene control sequences; and administering a cathepsin K polypeptide or polynucleotide to
25 an organism to augment cathepsin K function or remediate cathepsin K dysfunction.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided probes that hybridize specifically to human cathepsin K sequences.

In certain additional preferred embodiments of this aspect of the invention there are
30 provided antibodies against cathepsin K polypeptides. In certain particularly preferred embodiments in this regard, the antibodies are highly selective for human cathepsin K.

In accordance with another aspect of the present invention, there are provided cathepsin K agonists. Among preferred agonists are molecules that mimic cathepsin K, that bind to cathepsin K-binding molecules or receptor molecules, and that elicit or augment

cathepsin K-induced responses. Also among preferred agonists are molecules that interact with cathepsin K or cathepsin K polypeptides, or with other modulators of cathepsin K activities, and thereby potentiate or augment an effect of cathepsin K or more than one effect of cathepsin K.

5 In accordance with yet another aspect of the present invention, there are provided cathepsin K antagonists. Among preferred antagonists are those which mimic cathepsin K so as to bind to cathepsin K receptor or binding molecules but not elicit a cathepsin K-induced response or more than one cathepsin K-induced response. Also among preferred antagonists are molecules that bind to or interact with cathepsin K so as to inhibit an effect
10 of cathepsin K or more than one effect of cathepsin K.

The agonists and antagonists may be used to mimic, augment or inhibit the action of cathepsin K polypeptides. They may be used, for instance, to treat osteoporosis, Paget's disease, Gaucher's disease, CNS inflammation, Alzheimer's disease, hyperparathyroidism, bone degradation, metastatic tumors, and degradation of bone implants and bone prostheses,
15 particularly dental implants. Such antagonists may be particularly useful to treat osteoporosis, Paget's disease, Gaucher's disease, Alzheimer's disease, hyperparathyroidism, bone degradation, metastatic tumors, CNS inflammation, rheumatoid arthritis, osteoarthritis, periodontal disease and degradation of bone implants and bone prostheses, particularly dental implants.

20 In a further aspect of the invention there are provided compositions comprising a cathepsin K polynucleotide or a cathepsin K polypeptide for administration to cells *in vitro*, to cells *ex vivo* and to cells *in vivo*, or to a multicellular organism. In certain particularly preferred embodiments of this aspect of the invention, the compositions comprise a cathepsin K polynucleotide for expression of a cathepsin K polypeptide in a host organism
25 for treatment of disease. Particularly preferred in this regard is expression in a human patient for treatment of a dysfunction associated with aberrant endogenous activity of cathepsin K or to provide therapeutic.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill from the following description. It should be understood, however,
30 that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings depict certain embodiments of the invention. They are illustrative only and do not limit the invention otherwise disclosed herein.

Figure 1 shows the genomic nucleotide sequence of human cathepsin K. [SEQ ID NO: 1]. Exons are capitalized and underlined. Intron sequence is in lower case letters.

Figure 2 shows the nucleotide, exon-intron boundaries and deduced amino acid sequence of human cathepsin K.

Figure 3 (A - S) shows structural features of cathepsin K [SEQ ID NO: 2-19].

Figure 4 shows the intron-exon junctions.

Figure 5 shows the regions of similarity between amino acid sequences of cathepsin K, human cathepsins S, L, H, B, D, E, G and rabbit OC2 polypeptides.

Figure 6 shows the deduced amino acid sequence of human cathepsin K.

Figure 7 shows the PCR primers to amplify genomic DNA.

GLOSSARY

The following illustrative explanations are provided to facilitate understanding of certain terms used frequently herein, particularly in the examples. The explanations are provided as a convenience and are not limitative of the invention.

DIGESTION of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan.

For analytical purposes, typically, 1 µg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes.

Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers.

Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well known methods that are routine for those skilled in the art.

GENETIC ELEMENT generally means a polynucleotide comprising a region that encodes a polypeptide or a region that regulates transcription or translation or other processes important to expression of the polypeptide in a host cell, or a polynucleotide comprising both a region that encodes a polypeptide and a region operably linked thereto that regulates expression.

Genetic elements may be comprised within a vector that replicates as an episomal element; that is, as a molecule physically independent of the host cell genome. They may be comprised within mini-chromosomes, such as those that arise during amplification of transfected DNA by methotrexate selection in eukaryotic cells. Genetic elements also may be comprised within a host cell genome; not in their natural state but, rather, following manipulation such as isolation, cloning and introduction into a host cell in the form of purified DNA or in a vector, among others.

IDENTITY means the degree of sequence relatedness between two polypeptide or two polynucleotides sequences as determined by the identity of the match between two strings of such sequences. Identity can be readily calculated. While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Methods commonly employed to determine identity between two sequences include, but are not limited to disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipman, D., SIAM J. Applied Math., 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the two sequences tested. Such methods are codified in computer programs. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403 (1990)).

ISOLATED means altered "by the hand of man" from its natural state; i.e., that, if it occurs in nature, it has been changed or removed from its original environment, or both.

For example, a naturally occurring polynucleotide or a polypeptide naturally present in a living animal in its natural state is not "isolated," but the same polynucleotide or polypeptide separated from some or all of the coexisting materials of its natural is "isolated", as the term is employed herein.

- 5 As part of or following isolation, such polynucleotides can be joined to other polynucleotides, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated polynucleotides, alone or joined to other polynucleotides such as vectors, can be introduced into host cells, in culture or in whole organisms. Introduced into host cells in culture or in whole organisms, such
- 10 DNAs still would be isolated, as the term is used herein, because they would not be in their naturally occurring form or environment. Similarly, the polynucleotides and polypeptides may occur in a composition, such as a media formulations, solutions for introduction of polynucleotides or polypeptides, for example, into cells, compositions or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring
- 15 compositions, and, therein remain isolated polynucleotides or polypeptides within the meaning of that term as it is employed herein.

- LIGATION refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double stranded DNAs. Techniques for ligation are well known to the art and protocols for ligation are described in standard
- 20 laboratory manuals and references, such as, for instance, Sambrook et al., MOLECULAR CLONING, A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989) and Maniatis et al., pg. 146, as cited below.

- OLIGONUCLEOTIDE(S) refers to relatively short polynucleotides. Often the term refers to single-stranded deoxyribonucleotides, but it can refer as well to single-, double-, or
- 25 triple-stranded ribonucleotides, antisense polynucleotides, RNA:DNA hybrids and double-stranded DNAs, among others.

- Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods,
- 30 including in vitro recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms.

Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form

recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP.

5 The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

10 PLASMIDS generally are designated herein by a lower case letter *p* preceded and/or followed by capital letters and/or numbers, in accordance with standard naming conventions that are familiar to those of skill in the art.

Starting plasmids disclosed herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids by routine application of well known, published procedures. Many plasmids and other cloning and expression
15 vectors that can be used in accordance with the present invention are well known and readily available to those of skill in the art. Moreover, those of skill readily may construct any number of other plasmids suitable for use in the invention. The properties, construction and use of such plasmids, as well as other vectors, in the present invention will be readily apparent to those of skill from the present disclosure.

20 POLYNUCLEOTIDE(S) generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. Thus, for instance, polynucleotides as used herein refers to, among others, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and
25 double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions.

In addition, polynucleotide as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be
30 from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide.

As used herein, the term polynucleotide includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones

modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein.

5 It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells, *inter alia*.

10 POLYPEPTIDES, as used herein, includes all polypeptides as described below. The basic structure of polypeptides is well known and has been described in innumerable textbooks and other publications in the art. In this context, the term is used herein to refer to any peptide or protein comprising two or more amino acids joined to each other in a linear chain by peptide bonds. As used herein, the term refers to both short chains, which
15 also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types.

It will be appreciated that polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids, and that many
20 amino acids, including the terminal amino acids, may be modified in a given polypeptide, either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques which are well known to the art. Even the common modifications that occur naturally in polypeptides are too numerous to list exhaustively here, but they are well described in basic texts and in more detailed
25 monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. Among the known modifications which may be present in polypeptides of the present are, to name an illustrative few, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of
30 a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation,

racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

- Such modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications,
- 5 glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, F.,
- 10 *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Analysis for protein modifications and nonprotein cofactors*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, *Ann. N.Y. Acad. Sci.*
- 15 663: 48-62 (1992).

- It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by
- 20 human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well.

- Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage
- 25 of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

- 30 The modifications that occur in a polypeptide often will be a function of how it is made. For polypeptides made by expressing a cloned gene in a host, for instance, the nature and extent of the modifications in large part will be determined by the host cell posttranslational modification capacity and the modification signals present in the polypeptide amino acid sequence. For instance, as is well known, glycosylation often does

not occur in bacterial hosts such as *E. coli*. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to express efficiently
5 mammalian proteins having native patterns of glycosylation, *inter alia*. Similar considerations apply to other modifications.

It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications.

10 In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that are present in polypeptides synthesized by expressing a polynucleotide in a host cell.

VARIANT(S) of polynucleotides or polypeptides, as the term is used herein, are polynucleotides or polypeptides that differ from a reference polynucleotide or polypeptide,
15 respectively. Variants in this sense are described below and elsewhere in the present disclosure in greater detail.

(1) A polynucleotide that differs in nucleotide sequence from another, reference polynucleotide. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

20 As noted below, changes in the nucleotide sequence of the variant may be silent. That is, they may not alter the amino acids encoded by the polynucleotide. Where alterations are limited to silent changes of this type a variant will encode a polypeptide with the same amino acid sequence as the reference. Also as noted below, changes in the nucleotide sequence of the variant may alter the amino acid sequence of a polypeptide
25 encoded by the reference polynucleotide. Such nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below.

(2) A polypeptide that differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference and
30 the variant are closely similar overall and, in many region, identical.

A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

RECEPTOR MOLECULE, as used herein, refers to molecules which bind or interact specifically with cathepsin K polypeptides of the present invention, including not only classic receptors and enzymatic substrates, both of which are preferred, but also other molecules that specifically bind to or interact with polypeptides of the invention (which also
5 may be referred to as "binding molecules" and "interaction molecules," respectively and as "cathepsin K binding molecules" and "cathepsin K interaction molecules." These cathepsin K binding molecules also include, for example, cathepsin K substrate analogs. Binding between polypeptides of the invention and such molecules, including receptor or binding or interaction molecules may be exclusive to polypeptides of the invention, which is very
10 highly preferred, or it may be highly specific for polypeptides of the invention, which is highly preferred, or it may be highly specific to a group of proteins that includes polypeptides of the invention, which is preferred, or it may be specific to several groups of proteins at least one of which includes polypeptides of the invention.

Receptors also may be non-naturally occurring, such as antibodies and
15 antibody-derived reagents that bind specifically to polypeptides of the invention.

DESCRIPTION OF THE INVENTION

The present invention relates to novel cathepsin K polypeptides and polynucleotides, among other things, as described in greater detail below. In particular, the
20 invention relates to polypeptides and polynucleotides of a novel human cathepsin K, which is related by amino acid sequence homology to rabbit OC-2 and human cathepsin O cDNA. Tezuka, K., et al., J. Biol. Chem., 269:1106-1109, (1994). The invention relates especially to cathepsin K having the nucleotide sequences set out in Figure 1 [SEQ ID NO: 1], and to the cathepsin K nucleotide sequences of the gDNA in ATCC Deposit No. 98035, which is
25 herein referred to as "the deposited clone" or as the "gDNA of the deposited clone." It will be appreciated that the nucleotide sequences set out in Figure 1 [SEQ ID NO: 1] were obtained by sequencing the gDNA of the deposited clone, as more specifically set forth elsewhere herein. Hence, the sequence of the deposited clone is controlling as to any discrepancies between the two.

30

Polynucleotides

In accordance with one aspect of the present invention, there are provided isolated polynucleotides which encode the cathepsin K polypeptide having the deduced amino acid

sequence of Figure 2 [SEQ ID NO:20] (see also Figure 6 for the deduced amino acid sequence) or the cathepsin K polypeptide encoded by the gDNA in the deposited clone.

Using the information provided herein, such as the polynucleotide sequence set out in Figure 1 [SEQ ID NO: 1], a polynucleotide of the present invention encoding human cathepsin K polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning gDNAs using DNA from cells of a human as starting material. Illustrative of the invention, the polynucleotide set out in Figure 1 [SEQ ID NO: 1] was discovered in a human gDNA library as described in Example 1.

Human cathepsin K of the invention is structurally related to other proteins of the cathepsin family, as shown by the results of sequencing the gDNA encoding human cathepsin K in the deposited clone. The gDNA sequence thus obtained is set out in Figure 1 [SEQ ID NO: 1]. It contains a non-contiguous open reading frame encoding, after intron removal, but including all exons, a protein of about 329 amino acid residues.

Polynucleotides of the present invention may be in the form of RNA, such as mRNA or hnRNA, or in the form of DNA, including, for instance, cDNA and gDNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The DNA may be triple-stranded, double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The coding sequence which encodes the polypeptide may be identical to the exon sequence of the polynucleotide shown in Figure 1 [SEQ ID NO: 1] or that of the deposited clone. It also may be a polynucleotide with a different sequence, which, as a result of the redundancy (degeneracy) of the genetic code, encodes the polypeptide of the DNA of Figure 2 [SEQ ID NO:20] or of the deposited gDNA, including, but not limited to, splice variants transcribed from such gDNA.

Polynucleotides of the present invention which encode the polypeptide of Figure 1 [SEQ ID NO: 1] or the polypeptide encoded by the deposited gDNA may include, but are not limited to the coding sequence for the mature polypeptide, by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences, including for example, but not limited to introns and non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing -

including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities. Thus, for instance, the polypeptide may be fused to a marker sequence, such as a peptide, which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, such as the tag provided in the vector pQE-9, among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci., USA 86: 821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The HA tag corresponds to an epitope derived of influenza hemagglutinin protein, which has been described by Wilson et al., Cell 37: 767 (1984), for instance.

In accordance with the foregoing, the term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides which include a sequence encoding a polypeptide of the present invention, particularly the human cathepsin K having the amino acid sequence set out in Figure 2 [SEQ ID NO:20] or the amino acid sequence of the human cathepsin K encoded by the gDNA of the deposited clone. The term encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by introns) together with additional regions, that also may contain coding and/or non-coding sequences.

The present invention further relates to variants of the herein above described polynucleotides which encode for fragments, analogs and derivatives of the polypeptide having the deduced amino acid sequence of Figure 2 [SEQ ID NO:20] or the polypeptide encoded by the exons of the gDNA of the deposited clone, including, but not limited to, splice variants transcribed from such gDNA. A variant of the polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant or splice variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms. Such non-naturally occurring variants of the polynucleotide may be made by modifying splice acceptor, donor and/or branch sites, or by expressing the gDNA in cells where it is not naturally expressed, or cell extracts made from such cells.

Among variants in this regard are variants that differ from the aforementioned polynucleotides by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in

coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

Among the particularly preferred embodiments of the invention in this regard are polynucleotide sequence of cathepsin K set out in Figure 1 [SEQ ID NO: 1] or the
5 polynucleotide sequence of cathepsin K of the gDNA of the deposited clone; variants, analogs, derivatives and fragments thereof, and fragments of the variants, analogs and derivatives.

Further particularly preferred in this regard are polynucleotides encoding cathepsin K variants, analogs, derivatives and fragments, and variants, analogs and derivatives of the
10 fragments, which have the amino acid sequence of the cathepsin K polypeptide of Figure 2 [SEQ ID NO:20] or of the deposit in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the cathepsin K. Also, especially preferred in this regard are
15 conservative substitutions. Most highly preferred are polypeptides having the amino acid sequence of Figure 2 [SEQ ID NO:20] or of the deposit, without substitutions.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical to a polynucleotide encoding the cathepsin K polypeptide having the amino acid sequence set out in Figure 2 [SEQ ID NO:20], or variants, close homologs, derivatives
20 and analogs thereof, as described above, and polynucleotides which are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical to a polynucleotide encoding the cathepsin K polypeptide of the gDNA of the deposited clone and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical to the same are particularly
25 preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Still further preferred embodiments of the invention are polynucleotides comprising
30 cathepsin K intron polynucleotide sequences, particularly polynucleotides comprising intron 1 [SEQ ID NO: 4], 2 [SEQ ID NO: 6], 3 [SEQ ID NO: 8], 4 [SEQ ID NO: 10], 5 [SEQ ID NO: 12], 6 [SEQ ID NO: 14] or 7 [SEQ ID NO: 16], having the intron polynucleotide sequence set out in Figures 1 [SEQ ID NO: 1] and 3 [SEQ ID NO: 2-19], or variants, close homologs, derivatives and analogs thereof, as described above, and polynucleotides which

are complementary to such polynucleotides. Other preferred embodiments of the invention are polynucleotides comprising cathepsin K intron 1 [SEQ ID NO: 4], 2 [SEQ ID NO: 6], 3 [SEQ ID NO: 8], 4 [SEQ ID NO: 10], 5 [SEQ ID NO: 12], 6 [SEQ ID NO: 14] or 7 [SEQ ID NO: 16], operatively linked to the exon of a gene other than cathepsin K, or joining a

5 cathepsin K exon and an exon of another gene.

Still other preferred embodiments of the invention are polynucleotides comprising cathepsin K exon polynucleotide sequences, particularly polynucleotides comprising exon 1 [SEQ ID NO: 3], 2 [SEQ ID NO: 5], 3 [SEQ ID NO: 7], 4 [SEQ ID NO: 9], 5 [SEQ ID NO: 11], 6 [SEQ ID NO: 13], 7 [SEQ ID NO: 15] or 8 [SEQ ID NO: 17], having the exon

10 polynucleotide sequence set out in Figures 1 [SEQ ID NO: 1] and 3 [SEQ ID NO: 2-19], or variants, close homologs, derivatives and analogs thereof, as described above, and polynucleotides which are complementary to such polynucleotides. Other preferred embodiments of the invention are polynucleotides comprising cathepsin K exon 1 [SEQ ID NO: 3], 2 [SEQ ID NO: 5], 3 [SEQ ID NO: 7], 4 [SEQ ID NO: 9], 5 [SEQ ID NO: 11], 6

15 [SEQ ID NO: 13], 7 [SEQ ID NO: 15] or 8 [SEQ ID NO: 17], operatively linked to the intron of a gene other than cathepsin K.

More preferred embodiments of the invention are differentially spliced polynucleotides, particularly those comprising any one or more of the following exon-exon pairs: 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 2-4, 2-5, 2-6, 2-7, 2-8, 3-4, 3-5, 3-6,

20 3-7, 3-8, 4-5, 4-6, 4-7, 4-8, 5-7, 5-8, or 6-8. Particularly preferred embodiments of the invention are differentially spliced polynucleotides which encode polypeptides which function in cells, especially those which have a biological activity of cathepsin K, most especially those expressed in human cells.

Polynucleotides comprising exon-exon pairs may be a naturally occurring variant

25 such as a naturally occurring splice variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms. Such non-naturally occurring variants of the polynucleotide may be made by modifying splice acceptor, donor and/or branch sites, or by expressing the gDNA in cells where it is

30 not naturally expressed, or cell extracts made from such cells. Exon-exon pairs can be full, fused exons or can be fused fragments of exons with a splice junction present. Preferred exon-exon pairs comprising exon fragments may be made from at least two exons, one of which comprises an operable splice donor site and the other of which comprises an operable splice acceptor site and which both are operatively linked by an intron.

Particularly preferred embodiments in this respect, moreover, are polynucleotides which encode polypeptides which retain substantially the same biological function or activity as the mature polypeptide encoded by the cDNA of Figure 2 (SEQ ID NO:20) or the gDNA of the deposited clone.

5 The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the
10 sequences.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding cathepsin K and to isolate cDNA and genomic clones of other genes that
15 have a high sequence similarity to the human cathepsin K gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of the cathepsin K gene may be isolated by
20 screening using the known DNA sequence to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the present invention is then used to screen a library of human cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the present invention may be employed as
25 research reagents and materials for discovery of treatments and diagnostics to human disease, as further discussed herein relating to polynucleotide assays, *inter alia*.

The polynucleotides may encode a polypeptide which is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance).
30 Such sequences may play a role in processing of a protein from precursor to a mature form, may facilitate protein trafficking, may prolong or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

- 5 In sum, a polynucleotide of the present invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences which are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during
10 processing steps that produce active and mature forms of the polypeptide.

Deposited materials

- A deposit containing a human cathepsin K gDNA has been deposited with the American Type Culture Collection, as noted above. Also as noted above, the gDNA
15 deposit is referred to herein as "the deposited clone" or as "the gDNA of the deposited clone."

The deposited clone was deposited with the American Type Culture Collection, 12301 Park Lawn Drive, Rockville, Maryland 20852, USA, on April 26, 1996, and assigned ATCC Deposit No. 98035.

- 20 The deposited material is a P1 cosmid that contains the full length cathepsin K gDNA, referred to as "P1SacB2CatK/P129" upon deposit.

- The deposit has been made under the terms of the Budapest Treaty on the international recognition of the deposit of micro-organisms for purposes of patent procedure. The strain will be irrevocably and without restriction or condition released to
25 the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. section 112.

- The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of
30 any conflict with any description of sequences herein.

A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

Polypeptides

The present invention further relates to a human cathepsin K polypeptide which has the deduced amino acid sequence of Figure 2 [SEQ ID NO:20], which is encoded by an unspliced or differentially spliced hnRNA or mRNA transcribed from the sequence of Figure 1 [SEQ ID NO: 1], or which has the amino acid sequence encoded by the deposited clone. Also provided are polypeptides encoded by the cathepsin K gDNA comprising missense or nonsense mutations, or those polypeptides encoded by unspliced or partially spliced hnRNAs which still comprise at least one intron, particularly those polypeptides which are naturally found in cells, especially human cells. Frameshift mutations have been shown to be associated with disease (Hol, FA, et al. Journal of Medical Genetics, 1995, 32 (1), 52-56).

Preferred polypeptides provided by the invention are encoded by differentially spliced polynucleotides, particularly those polypeptides encoded by polynucleotides comprising any one or more of the following exon-exon pairs: 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 2-4, 2-5, 2-6, 2-7, 2-8, 3-4, 3-5, 3-6, 3-7, 3-8, 4-5, 4-6, 4-7, 4-8, 5-7, 5-8, or 6-8. Particularly preferred embodiments of the invention are polypeptides encoded by differentially spliced polynucleotides, which polypeptides function in cells, especially those which have a biological activity of cathepsin K, most especially those expressed in human cells.

Still further preferred embodiments of the invention are polypeptides encoded by polynucleotides comprising exon polynucleotide sequences, particularly polynucleotides comprising cathepsin K exon 1 [SEQ ID NO: 3], 2 [SEQ ID NO: 5], 3 [SEQ ID NO: 7], 4 [SEQ ID NO: 9], 5 [SEQ ID NO: 11], 6 [SEQ ID NO: 13], 7 [SEQ ID NO: 15] or 8 [SEQ ID NO: 17], having the exon polynucleotide sequence set out in Figures 1 [SEQ ID NO: 1] and 3 [SEQ ID NO: 2-19], or variants, close homologs, derivatives and analogs thereof, as described above, and polypeptides encoded by polynucleotides which are complementary to such polynucleotides. Other preferred embodiments of the invention are polypeptides encoded by polynucleotides comprising comprising cathepsin K exon exon 1 [SEQ ID NO: 3], 2 [SEQ ID NO: 5], 3 [SEQ ID NO: 7], 4 [SEQ ID NO: 9], 5 [SEQ ID NO: 11], 6 [SEQ ID NO: 13], 7 [SEQ ID NO: 15] or 8 [SEQ ID NO: 17], operatively linked to the intron of a gene other than cathepsin K, or joined to an exon of another gene.

The invention also relates to fragments, analogs and derivatives of these polypeptides. The terms "fragment," "derivative" and "analog" when referring to the polypeptide of Figure 2 [SEQ ID NO:20], a polypeptide encoded by an unspliced or

differentially spliced hnRNA or mRNA transcribed from the sequence of Figure 1 [SEQ ID NO: 1], or that encoded by the deposited gDNA, means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce
5 an active mature polypeptide.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide. In certain preferred embodiments it is a recombinant polypeptide.

The fragment, derivative or analog of the polypeptide of Figure 2 [SEQ ID NO:20],
10 or that encoded by an unspliced or differentially spliced hnRNA or mRNA transcribed from the sequence of Figure 1 [SEQ ID NO: 1], or that encoded by the gDNA in the deposited clone may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic
15 code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the
20 mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Among the particularly preferred embodiments of the invention in this regard are polypeptides having the amino acid sequence of cathepsin K set out in Figure 2 [SEQ ID NO:20], variants, analogs, derivatives and fragments thereof, and variants, analogs and
25 derivatives of the fragments. Alternatively, particularly preferred embodiments of the invention in this regard are polypeptides having the amino acid sequence of the cathepsin K of the gDNA in the deposited clone, variants, analogs, derivatives and fragments thereof, and variants, analogs and derivatives of the fragments.

Among preferred variants are those that vary from a reference by conservative
30 amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic

residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Further particularly preferred in this regard are variants, analogs, derivatives and fragments, and variants, analogs and derivatives of the fragments, having the amino acid sequence of the cathepsin K polypeptide of Figure 2 (SEQ ID NO:20) or of the gDNA in the deposited clone, in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the cathepsin K. Also especially preferred in this regard are conservative substitutions. Most highly preferred are polypeptides having the amino acid sequence of Figure 2 (SEQ ID NO:20) or the deposited clone without substitutions.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The polypeptides of the present invention include the polypeptide encoded by at least one of the exons of SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15 or SEQ ID NO: 17, (in particular the mature polypeptide) as well as polypeptides which have at least 70% similarity (preferably at least 70% identity) to the polypeptide encoded by at least one of the exons of SEQ ID NO: SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15 or SEQ ID NO: 17 and more preferably at least 90% similarity (more preferably at least 90% identity) to the polypeptide encoded by at least one of the exons of SEQ ID NO: SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15 or SEQ ID NO: 17 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the polypeptide encoded by at least one of the exons of SEQ ID NO: SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15 or SEQ ID NO: 17 and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide.

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis;

therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

Fragments

5 Also among preferred embodiments of this aspect of the present invention are polypeptides comprising fragments of cathepsin K, most particularly fragments of the cathepsin K having the amino acids encoded by the exons set out in Figure 1 [SEQ ID NO: 1], or having the amino acid encoded by the exon sequence of the cathepsin K of the deposited clone, and exon fragments or variants and derivatives of the cathepsin K of Figure
10 1 [SEQ ID NO: 1] or of the deposited clone.

In this regard a fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned cathepsin K polypeptides and variants or derivatives thereof.

Such fragments may be "free-standing," i.e., not part of or fused to other amino
15 acids or polypeptides, such as, for example, an exon, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the presently discussed fragments most preferably form a single continuous region. However, several fragments may be comprised within a single larger polypeptide. For instance, certain preferred embodiments relate to a fragment of a cathepsin K
20 polypeptide of the present comprised within a precursor polypeptide designed for expression in a host and having heterologous pre and pro-polypeptide regions fused to the amino terminus of the cathepsin K fragment and an additional region fused to the carboxyl terminus of the fragment. Therefore, fragments in one aspect of the meaning intended herein, refers to the portion or portions of a fusion polypeptide or fusion protein derived
25 from cathepsin K.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which are encoded by the polynucleotide sequence comprising cathepsin K exon 1, 2, 3, 4, 5, 6, 7 or 8, having the exon or intron 1,2,3,4,5,6 or 7 polynucleotide sequences respectively as set out in Figures 1
30 [SEQ ID NO: 1] and 3 [SEQ ID NO: 2-19], or variants, close homologs, derivatives and analogs thereof, as described above, and polypeptides encoded by polynucleotides which are complementary to such polynucleotides.

In this context about includes the particularly recited range and ranges larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

For instance, about 65-90 amino acids in this context means a polypeptide fragment of 65 plus or minus several, a few, 5, 4, 3, 2 or 1 amino acids to 90 plus or minus several a few, 5, 4, 3, 2 or 1 amino acid residues, i.e., ranges as broad as 65 minus several amino acids to 90 plus several amino acids to as narrow as 65 plus several amino acids to 90 minus several amino acids.

Highly preferred in this regard are the recited ranges plus or minus as many as 5 amino acids at either or at both extremes. Particularly highly preferred are the recited ranges plus or minus as many as 3 amino acids at either or at both the recited extremes. Especially particularly highly preferred are ranges plus or minus 1 amino acid at either or at both extremes or the recited ranges with no additions or deletions. Most highly preferred of all in this regard are fragments encoded by each of the exons of cathepsin K.

Among especially preferred fragments of the invention are truncation mutants of cathepsin K. Truncation mutants include cathepsin K polypeptides having the amino acid sequence encoded by the exons of Figure 1 [SEQ ID NO: 1], or of the deposited clone, or of variants or derivatives thereof, except for deletion of a continuous series of residues (that is, a continuous region, part or portion) that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or, as in double truncation mutants, deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Fragments having the size ranges set out about also are preferred embodiments of truncation fragments, which are especially preferred among fragments generally.

Also preferred in this aspect of the invention are fragments characterized by structural or functional attributes of cathepsin K. Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions ("alpha-regions"), beta-sheet and beta-sheet-forming regions ("beta-regions"), turn and turn-forming regions ("turn-regions"), coil and coil-forming regions ("coil-regions"), hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions and high antigenic index regions of cathepsin K.

Certain preferred regions include Garnier-Robson alpha-regions, beta-regions, turn-regions and coil-regions, Chou-Fasman alpha-regions, beta-regions and turn-regions, Kyte-Doolittle hydrophilic regions and hydrophilic regions, Eisenberg alpha and beta amphipathic regions, Karplus-Schulz flexible regions, Emmini surface-forming regions and Jameson-Wolf high antigenic index regions.

Among highly preferred fragments in this regard are those that comprise regions of cathepsin K that combine several structural features, such as several of the features set out above. In this regard, the exon sequences of Figure 1

[SEQ ID NO: 1], which all are characterized by encoding amino acid compositions highly characteristic of turn-regions, hydrophilic regions, flexible-regions, surface-forming regions, and high antigenic index-regions, are especially highly preferred regions. Such regions may be comprised within a larger polypeptide or may be by themselves a preferred fragment of the present invention, as discussed above. It will be appreciated that the term "about" as used in this paragraph has the meaning set out above regarding fragments in general.

Further preferred regions are those that mediate activities of cathepsin K. Most highly preferred in this regard are fragments that have a chemical, biological, antigenic or other activity of cathepsin K, including those with a similar activity or an improved activity, or with a decreased undesirable activity.

It will be appreciated that the invention also relates to, among others, polynucleotides encoding the aforementioned fragments, polynucleotides that hybridize to polynucleotides encoding the fragments, particularly those that hybridize under stringent conditions, and polynucleotides, such as PCR primers, for amplifying polynucleotides that encode the fragments. In these regards, preferred polynucleotides are those that correspondent to the preferred fragments, as discussed above.

Other preferred polynucleotides are genetic elements of cathepsin K, including, but not being limited to, a polyadenylation region, enhancers, a promoter, a cap site introns, exons, and splice sites (references describing these elements include, Darnel, J. et al. *Molecular Cell Biology*, second edition, W.H. Freeman, New York (1990); Watson, J.D., et al. *Molecular Biology of the Gene*, Benjamin/Cummings Pub., Menlo Park, CA, (1987)).

Untranslated regions contain many elements important in regulating gene expression. Mutations and markers in these regions have also been associated with disease (Ozawa T, et al., *European Journal of Immunogenetics*, APR 1995, 22 (2), 163-169). A preferred embodiment of the invention is the 5'UTR, particularly the sequence set forth in Figure 3(A) [SEQUENCE ID NO: 2]. Mutations and markers in the 5' UTR have been associated with disease (Carlock L, et al., *Human Genetics*, APR 1994, 93 (4), 457-459). A particularly preferred polynucleotide is an enhancer and promoter in the 5' UTR region of the cathepsin K gDNA. Enhancers are often found in the 5' UTR and upregulate gene expression (see Miller et al., *Biotechniques* 7: 980-990 (1989) for a general reference on promoters). The enhancer of the present invention can be operatively fused to heterologous

- genes to upregulate gene expression. It is believed that the enhancer promoter will regulate tissue-specific gene expression, being particularly useful to express genes in osteoclast and leukocytes, particularly macrophages cells. A particularly preferred polynucleotide is the enhancer promoter having the sequence set forth in Figure 3(A) [SEQUENCE ID NO: 2].
- 5 Transcription factors are often associated with the enhancer and promoter and act to modulate the function of these regions and binding sites for these factors have been described (Faisst, Steffen and Meyer, Silke, *Nucleic Acids Research*, Vol. 20, No. 1, pp. 3-26, 1991; Smale, Stephen T., *Transcription: Mechanisms and Regulation*, Raven Press, Ltd. pp. 63-81 (1994)). These sites bind such factors as, for example, Sp1, Ap1, and Ap3 which
- 10 are involved in transcription initiation (Faisst, Steffen and Meyer, Silke, *Nucleic Acids Research*, Vol. 20, No. 1, pp. 3-26, 1991). Preferred canonical binding sites for transcription factors are underlined in Figure 3(S) [SEQUENCE ID NO: 2]. The Pu Box in Figure 3(S) [SEQUENCE ID NO: 2] has been described to be present in a macrophage gene, a cell in which cathepsin K is also found (Zhang, Dong-Er, *Mol. and Cell. Biol.*, Vol.
- 15 14, No. 1, pp. 373-381 (1994)). The present invention provides a promoter region that is useful, among other things, for the mediation of tissue-specific expression in osteoclasts and leukocytes, particularly macrophages. A Pu box (AGGAA), present in the enhancer and promoter region has also been observed in a macrophage cell line (THP1). Pu boxes in the sequences in the invention are provided. These Pu boxes are believed to be active in the
- 20 cathepsin K gene in macrophages. RT-PCR performed in THP1 cells, using cathepsin K sequence as a probe, showed expression. The promoter is particularly useful for the study of the control of cathepsin K gene expression, particularly as a region to be probed to diagnose disease. Vitamin D response elements have been found in the 5'UTR of known genes (Kahlen, Jean-Pierre and Carlberg, Carsten, *Biochemical & Biophysical Research Communications*, Vol. 202, No. 3, pp. 1366-1372 (1994); Darwish, Hisham and DeLuca, Hector, *Critical Reviews in Eukaryotic Gene Expression*, 3(2):89-116 (1993); Carlberg, Carsten, *Eur. J. Biochem.* 231, pp. 517-527 (1995); Ohyama, Yoshihiko, *J. Biol. Chem.*, Vol. 269, No. 14, pp. 10545-10550 (1994)). Portions of vitamin D ("vD half sites") responsive elements and calcium ion responsive elements ("Ca half pairs") are present in
- 25 the 3' UTR sequence as set forth in Figure 3(S) [SEQUENCE ID NO: 2]. Such sites have been described (Katz, Ronald, W., Subauste, Jose, S., and Koenig, Ronald J., *J. Biol. Chem.*, Vol. 270, No. 10, pp. 5238-5242 (1995)). Other half sites present in the sequence of the 5'UTR set forth in Figure 3(A and S) [SEQUENCE ID NO: 2] include osteopontin/parathyroid hormone responsive element, calcitriol response element and
- 30

osteocalcin half site (see, for example, Juge-Aubry, Cristiana, et al., J. Biol. Chem., Vol. 270, No. 30, pp. 18117-18122 (1995)). Promoter factor binding sites found in the promoter and enhancer region and provided in the invention are also found in cathepsin K introns. Estrogen response elements are also expected to be present in cathepsin K 5' UTR. Skilled
5 artisans can readily find such elements using the methods provided herein.

A further preferred polynucleotide is a cap site located 49 base pairs upstream of the ATG start codon of the sequence set forth in (Figure 2).

A further preferred embodiment of the invention is the promoter region of cathepsin K (Figure 3(A) and (S) [SEQUENCE ID NO: 2]). Functional promoter region sequences
10 have been described (Corden, J., et al., Science, 209, pp. 1406-1414 (1990)). A non-canonical promoter region in the sequence of cathepsin K set forth in Figures 3(A) [SEQUENCE ID NO: 2] and (S) [SEQUENCE ID NO: 2] comprises an A-T rich stretch at 19-27 base pairs upstream of the start codon ATG. Mutations in the TATA box region of promoters have been shown to be associated with disease (Peltoketo H, et al., Genomics,
15 1994, 23 (1), 250-252).

The 3' untranslated region of cathepsin K is a preferred polynucleotide of the invention, especially that polynucleotide set forth in Figure 3(Q) [SEQUENCE ID NO: 18], especially that region set forth in Figure 3(R) [SEQUENCE ID NO: 19]. Mutations in the 3' UTR have been associated with disease (Saito A, et al., Journal of the American Society of
20 Nephrology, 1994, 4 (9), 1649-1653; Payne SJ, et al., Human Molecular Genetics, 1994, 3 (2), 390). The polyadenylation region set forth in Figure 3(Q) [SEQUENCE ID NO: 18] is also a preferred polynucleotide of the 3' UTR. The polyadenylation region comprises two copies of the canonical polyadenylation hexanucleotide, AATAAA. The polyadenylation region can be used, for example, in expression vectors to mediate mRNA 3' end formation
25 (see, for example Gil, A. et al., Nature 312:473-474 (London) (1984)).

Other particularly preferred polynucleotides of the invention are the splice sites, including, but not limited to the splice donors, splice acceptors and the splice branchpoint. Splice junctions formation is essential for the proper creation of an open reading frame (Mount. Stephen, M., Department of Molecular Biophysics and Biochemistry, Yale
30 University, Sterling Hall of Medicine, New Haven, CT, USA, IRL Press Limited, London, pp. 459-472 (1981)). Diseases associated with the improper formation of the splice junction are known. Particularly preferred splice junction polynucleotides are set forth in Figure 4.

Introns comprise elements important in gene expression and in the formation of mature mRNA. Mutations and markers in introns have been shown to be associated with diseases (Peral, G. et al., Human Molecular Genetics, APR 1995, 4 (4), 569-574; Chrysogelos, S.A., Nucleic Acids Research, 1993, 21 (24), 5736-5741; Ameis, D., Journal of Lipid Research, 1995, 36 (2), 241-250). The splice junctions have also been shown to be associated with disease (Ameis D, et al., Journal of Lipid Research, FEB 1995, 36 (2), 241-250; Petrini JHJ, et al., Journal of Immunology, 1994, 152 (1), 176-183; Kleiman FE, et al., Human Genetics, 1994, 94 (3), 279-282). Alternative splicing and cryptic splice sites selection also have been shown to be associated with disease (Arakawa H, et al., Human Molecular Genetics, 1994, 3 (4), 565-568; Tieu PT, et al., Human Mutation, 1994, 3 (3), 333-336; Reale MA, et al., Cancer Research, 1994, 54 (16), 4493-4501). Introns may also comprise enhancer elements as part of their sequence.

Preferred embodiments of the invention are the cathepsin K introns, particularly those introns having the sequences set forth in Figure 3 (C, E, G, I, K, M, and O) [SEQUENCE ID NO: 4, 6, 8, 10, 14, and 16]. Polymorphisms in the introns can serve as markers for disease following linkage analysis. Moreover, genetic analyses described herein can be used to locate mutations in the introns associated with and/or causing disease.

Another preferred embodiment is a cathepsin K intronic enhancer.

Intron 3 does not follow consensus splice junction GT/AG rule. This intron/exon boundaries was verified by sequencing of the P1 clone and the genomic DNA. GC/AG splice junctions though not common, have been described (Mount, Stephen, M., Department of Molecular Biophysics and Biochemistry, Yale University, Sterling Hall of Medicine, New Haven, CT, USA, IRL Press Limited, London, pp. 459-472 (1981)).

Further preferred embodiments of the invention are the cathepsin K exons, particularly those exons having the sequences set forth in Figure 3 (B, D, F, H, J, L, N, and P) [SEQUENCE ID NO: 3, 5, 7, 9, 11, 13, 15 and 17 respectively]. Polymorphisms in the exons can serve as markers for disease following linkage analysis. Moreover, genetic analyses described herein can be used to locate mutations in the exons associated with and/or causing disease.

Polynucleotide fragments of the invention can be used to create ribozymes that inhibit the expression of the cathepsin K gene. General methods for the construction of ribozyme constructs are known in the art (Stram Y, and Molad T, Virus Genes, 1995, 9 (2), 155-159). Skilled artisans can readily adapt these methods using the novel fragments of the invention to create novel ribozyme constructs. Preferred ribozyme constructs comprise

Introns comprise elements important in gene expression and in the formation of mature mRNA. Mutations and markers in introns have been shown to be associated with diseases (Peral, G. et al., Human Molecular Genetics, APR 1995, 4 (4), 569-574; Chrysogelos, S.A., Nucleic Acids Research, 1993, 21 (24), 5736-5741; Arneis, D., Journal of Lipid Research, 1995, 36 (2), 241-250). The splice junctions have also been shown to be associated with disease (Arneis D, et al., Journal of Lipid Research, FEB 1995, 36 (2), 241-250; Petrini JHJ, et al., Journal of Immunology, 1994, 152 (1), 176-183; Kleiman FE, et al., Human Genetics, 1994, 94 (3), 279-282). Alternative splicing and cryptic splice sites selection also have been shown to be associated with disease (Arakawa H, et al., Human Molecular Genetics, 1994, 3 (4), 565-568; Tieu PT, et al., Human Mutation, 1994, 3 (3), 333-336; Reale MA, et al., Cancer Research, 1994, 54 (16), 4493-4501). Introns may also comprise enhancer elements as part of their sequence.

Preferred embodiments of the invention are the cathepsin K introns, particularly those introns having the sequences set forth in Figure 3 (C, E, G, I, K, M, and O) [SEQUENCE ID NO: 4, 6, 8, 10, 14, and 16]. Polymorphisms in the introns can serve as markers for disease following linkage analysis. Moreover, genetic analyses described herein can be used to locate mutations in the introns associated with and/or causing disease.

Another preferred embodiment is a cathepsin K intronic enhancer.

Intron 3 does not follow consensus splice junction GT/AG rule. This intron/exon boundaries was verified by sequencing of the P1 clone and the genomic DNA. GC/AG splice junctions though not common, have been described (Mount, Stephen, M., Department of Molecular Biophysics and Biochemistry, Yale University, Sterling Hall of Medicine, New Haven, CT, USA, IRL Press Limited, London, pp. 459-472 (1981)).

Further preferred embodiments of the invention are the cathepsin K exons, particularly those exons having the sequences set forth in Figure 3 (B, D, F, H, J, L, N, and P) [SEQUENCE ID NO: 3, 5, 7, 9, 11, 13, 15 and 17 respectively]. Polymorphisms in the exons can serve as markers for disease following linkage analysis. Moreover, genetic analyses described herein can be used to locate mutations in the exons associated with and/or causing disease.

Polynucleotide fragments of the invention can be used to create ribozymes that inhibit the expression of the cathepsin K gene. General methods for the construction of ribozyme constructs are known in the art (Stram Y, and Molad T, Virus Genes, 1995, 9 (2), 155-159). Skilled artisans can readily adapt these methods using the novel fragments of the invention to create novel ribozyme constructs. Preferred ribozyme constructs comprise

sequences which are complementary to the transcribed control elements of the cathepsin K gene, particularly polynucleotides that are complementary to the 5' untranslated region, splice junctions, and 3' untranslated region, especially the polyadenylation region.

5 The fragments of the invention, particularly regions in the untranslated region, the promoter and introns are useful as diagnostic probes for disease, particularly bone disease, such as osteoporosis, and including, for example, Paget's disease, Gaucher's disease, CNS inflammation, Alzheimer's disease, hyperparathyroidism, bone degradation, metastatic tumors, rheumatoid arthritis, osteoarthritis, periodontal disease and degradation of bone implants and bone prostheses, particularly dental implants. Moreover, markers for disease
10 can be located in regions of the cathepsin gene, particularly untranslated regions, which are useful with the diagnostic methods of the invention.

Vectors, host cells, expression

15 The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells can be genetically engineered to incorporate polynucleotides and express polypeptides of the present invention. For instance, polynucleotides may be introduced into host cells using well known techniques of infection, transduction, transfection, transvection
20 and transformation. The polynucleotides may be introduced alone or with other polynucleotides. Such other polynucleotides may be introduced independently, co-introduced or introduced joined to the polynucleotides of the invention.

Thus, for instance, polynucleotides of the invention may be transfected into host cells with another, separate, polynucleotide encoding a selectable marker, using standard
25 techniques for co-transfection and selection in, for instance, mammalian cells. In this case the polynucleotides generally will be stably incorporated into the host cell genome.

Alternatively, the polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. The vector construct may be introduced into host cells by the aforementioned techniques. Generally, a plasmid vector is introduced as DNA in a
30 precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. Electroporation also may be used to introduce polynucleotides into a host. If the vector is a virus, it may be packaged in vitro or introduced into a packaging cell and the packaged virus may be transduced into cells. A wide variety of techniques suitable for making polynucleotides and for introducing polynucleotides into cells in accordance with this

aspect of the invention are well known and routine to those of skill in the art. Such techniques are reviewed at length in Sambrook et al. cited above, which is illustrative of the many laboratory manuals that detail these techniques.

5 In accordance with this aspect of the invention the vector may be, for example, a plasmid vector, a single or double-stranded phage vector, a single or double-stranded RNA or DNA viral vector. Such vectors may be introduced into cells as polynucleotides, preferably DNA, by well known techniques for introducing DNA and RNA into cells. The vectors, in the case of phage and viral vectors also may be and preferably are introduced into cells as packaged or encapsidated virus by well known techniques for infection and
10 transduction. Viral vectors may be replication competent or replication defective. In the latter case viral propagation generally will occur only in complementing host cells.

Preferred among vectors, in certain respects, are those for expression of polynucleotides and polypeptides of the present invention. Generally, such vectors comprise cis-acting control regions effective for expression in a host operatively linked to
15 the polynucleotide to be expressed. Appropriate trans-acting factors either are supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression. Such specific expression may be inducible expression or expression only in
20 certain types of cells or both inducible and cell-specific. Particularly preferred among inducible vectors are vectors that can be induced for expression by environmental factors that are easy to manipulate, such as temperature and nutrient additives. A variety of vectors suitable to this aspect of the invention, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic hosts, are well known and employed routinely
25 by those of skill in the art.

The engineered host cells can be cultured in conventional nutrient media, which may be modified as appropriate for, *inter alia*, activating promoters, selecting transformants or amplifying genes. Culture conditions, such as temperature, pH and the like, previously used with the host cell selected for expression generally will be suitable for expression of
30 polypeptides of the present invention as will be apparent to those of skill in the art.

A great variety of expression vectors can be used to express a polypeptide of the invention. Such vectors include chromosomal, episomal and virus-derived vectors e.g., vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as

SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids, all may be used for expression in accordance with this aspect of the present invention. Generally, any
5 vector suitable to maintain, propagate or express polynucleotides to express a polypeptide in a host may be used for expression in this regard.

The appropriate DNA sequence may be inserted into the vector by any of a variety of well-known and routine techniques. In general, a DNA sequence for expression is joined to an expression vector by cleaving the DNA sequence and the expression vector with one
10 or more restriction endonucleases and then joining the restriction fragments together using T4 DNA ligase. Procedures for restriction and ligation that can be used to this end are well known and routine to those of skill. Suitable procedures in this regard, and for constructing expression vectors using alternative techniques, which also are well known and routine to those skill, are set forth in great detail in Sambrook et al. cited elsewhere herein.

15 The DNA sequence in the expression vector is operatively linked to appropriate expression control sequence(s), including, for instance, a promoter to direct mRNA transcription. Representatives of such promoters include the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name just a few of the well-known promoters. It will be understood that
20 numerous promoters not mentioned are suitable for use in this aspect of the invention are well known and readily may be employed by those of skill in the manner illustrated by the discussion and the examples herein.

In general, expression constructs will contain sites for transcription initiation and termination, and, in the transcribed region, a ribosome binding site for translation. The
25 coding portion of the mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

In addition, the constructs may contain control regions that regulate as well as engender expression. Generally, in accordance with many commonly practiced procedures,
30 such regions will operate by controlling transcription, such as repressor binding sites and enhancers, among others.

Vectors for propagation and expression generally will include selectable markers. Such markers also may be suitable for amplification or the vectors may contain additional markers for this purpose. In this regard, the expression vectors preferably contain one or

more selectable marker genes to provide a phenotypic trait for selection of transformed host cells. Preferred markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, tetracycline, kanamycin, and ampicillin resistance genes for culturing *E. coli* and other bacteria.

5 The vector containing the appropriate DNA sequence as described elsewhere herein, as well as an appropriate promoter, and other appropriate control sequences, may be introduced into an appropriate host using a variety of well known techniques suitable to expression therein of a desired polypeptide. Representative examples of appropriate hosts include bacterial cells, such as *E. coli*, streptomyces and *Salmonella typhimurium* cells;
10 fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Hosts for a great variety of expression constructs are well known, and those of skill will be enabled by the present disclosure readily to select a host for expressing a polypeptides in accordance with this aspect of the present invention.

15 More particularly, the present invention also includes recombinant constructs, such as expression constructs, comprising one or more of the sequences described above. The constructs comprise a vector, such as a plasmid or viral vector, into which such a sequence of the invention has been inserted. The sequence may be inserted in a forward or reverse orientation. In certain preferred embodiments in this regard, the construct further comprises
20 regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and there are many commercially available vectors suitable for use in the present invention.

 The following vectors, which are commercially available, are provided by way of example. Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9,
25 available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and
30 pSVK3, pBPV, pMSG and pSVL available from Pharmacia. These vectors are listed solely by way of illustration of the many commercially available and well known vectors that are available to those of skill in the art for use in accordance with this aspect of the present invention. It will be appreciated that any other plasmid or vector suitable for, for example, introduction, maintenance, propagation or expression of a polynucleotide or polypeptide of the invention in a host may be used in this aspect of the invention.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit lacking a promoter region, such as a chloramphenicol acetyl transferase ("CAT") transcription unit, downstream of restriction site or sites for introducing a candidate promoter fragment; i.e., a fragment that may contain a promoter.

- 5 As is well known, introduction into the vector of a promoter-containing fragment at the restriction site upstream of the cat gene engenders production of CAT activity, which can be detected by standard CAT assays. Vectors suitable to this end are well known and readily available. Two such vectors are pKK232-8 and pCM7. Thus, promoters for expression of polynucleotides of the present invention include not only well known and
- 10 readily available promoters, but also promoters that readily may be obtained by the foregoing technique, using a reporter gene.

- A preferred embodiment of the invention are expression vectors comprising cathepsin K promoter sequences that function as a promoter. Such vector constructs may be used for targeted gene expression in cells which utilize the cathepsin K promoter, for
- 15 example, osteoclasts and macrophages. Any gene of interest can be expressibly linked to the cathepsin K promoter and expressed in such cells which utilize the cathepsin K promoter. In this manner genes which immortalize primary eukaryotic cells, such as, for example, SV40 T-Antigen, may be expressibly linked cathepsin K promoter to immortalize cells, such as, for example, bone cells, including osteoclasts, and macrophages. Certain
- 20 preferred vectors comprise cathepsin K promoter expressibly linked to a toxin gene, such as for example, ricin, and are useful in methods for the targeted killing of cell populations that utilize the cathepsin K promoter for gene expression. Certain other preferred vectors comprise cathepsin K promoter expressibly linked to a anti-cathepsin K ribozyme or antisense polynucleotide, which are useful in methods for such targeted killing.

- 25 Among known bacterial promoters suitable for expression of polynucleotides and polypeptides in accordance with the present invention are the *E. coli* lacI and lacZ promoters, the T3 and T7 promoters, the gpt promoter, the lambda PR, PL promoters and the trp promoter. Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40
- 30 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Selection of appropriate vectors and promoters for expression in a host cell is a well known procedure and the requisite techniques for expression vector construction,

introduction of the vector into the host and expression in the host are routine skills in the art.

5 The present invention also relates to host cells containing the above-described constructs discussed above. The host cell can be a higher eukaryotic cell, such as a mammalian cell or insect cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell.

10 Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al. BASIC METHODS IN MOLECULAR BIOLOGY, (1986).

Constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

15 Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

25 Generally, recombinant expression vectors for yeast will include origins of replication, a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence, and a selectable marker to permit isolation of vector containing cells after exposure to the vector. Among suitable promoters are those derived from the genes that encode glycolytic enzymes such as 3-phosphoglycerate kinase ("PGK"), a-factor, acid phosphatase, and heat shock proteins, among others. Selectable markers include the ampicillin resistance gene of *E. coli* and the *trp1* gene of *S. cerevisiae*.

30 Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are *cis*-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication

origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Polynucleotides of the invention, encoding the heterologous structural sequence of a polypeptide of the invention generally will be inserted into the vector using standard techniques so that it is operably linked to the promoter for expression. The polynucleotide will be positioned so that the transcription start site is located appropriately 5' to a ribosome binding site. The ribosome binding site will be 5' to the AUG that initiates translation of the polypeptide to be expressed. Generally, there will be no other open reading frames that begin with an initiation codon, usually AUG, and lie between the ribosome binding site and the initiating AUG. Also, generally, there will be a translation stop codon at the end of the polypeptide and there will be a polyadenylation signal and a transcription termination signal appropriately disposed at the 3' end of the transcribed region.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, a region may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art.

Suitable prokaryotic hosts for propagation, maintenance or expression of polynucleotides and polypeptides in accordance with the invention include *Escherichia coli*, *Bacillus subtilis* and *Salmonella typhimurium*. Various species of *Pseudomonas*, *Streptomyces*, and *Staphylococcus* are suitable hosts in this regard. Moreover, many other hosts also known to those of skill may be employed in this regard.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example,

pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

5 Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, where the selected promoter is inducible it is induced by appropriate means (e.g., temperature shift or exposure to chemical inducer) and cells are cultured for an additional period.

Cells typically then are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

10 Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can be employed for expression, as well. Examples of mammalian expression systems include the COS-7 lines of monkey kidney
15 fibroblast, described in Gluzman et al., Cell 23: 175 (1981). Other cell lines capable of expressing a compatible vector include for example, the C127, 3T3, CHO, HeLa, human kidney 293 and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation
20 sites, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences that are necessary for expression. In certain preferred embodiments in this regard DNA sequences derived from the SV40 splice sites, and the SV40 polyadenylation sites are used for required non-transcribed genetic elements of these types.

25 The cathepsin K polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid
30 chromatography ("HPLC") is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a

prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial
5 modified methionine residue, in some cases as a result of host-mediated processes.

Further illustrative aspects and preferred embodiments of the invention

Cathepsin K polynucleotides and polypeptides may be used in accordance with the present invention for a variety of applications, particularly those that make use of the chemical and biological properties cathepsin K. Among these are applications in the
10 detection and treatment of disease, particularly bone disease, such as osteoporosis, and including, for example, Paget's disease, Gaucher's disease, CNS inflammation, Alzheimer's disease, hyperparathyroidism, bone degradation, metastatic tumors, rheumatoid arthritis, osteoarthritis, periodontal disease and degradation of bone implants and bone prostheses, particularly dental implants. Additional applications relate to diagnosis and to treatment of
15 disorders of cells, tissues and organisms. These aspects of the invention are illustrated further by the following discussion.

Polynucleotide assays

This invention is also related to the use of the cathepsin K exons, introns, promoters
20 and polynucleotides to detect complementary polynucleotides such as, for example, as a diagnostic reagent. Detection of a mutated form of cathepsin K associated with a dysfunction will provide a diagnostic tool that can add or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of cathepsin K, such as, for example, osteoporosis, periodontal disease, Paget's
25 disease, Gaucher's disease, CNS inflammation, Alzheimer's disease, hyperparathyroidism, and bone degradation, metastatic tumors, and degradation of bone implants and bone prostheses, particularly dental implants.

Individuals carrying mutations in the human cathepsin K gene may be detected at the DNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from
30 a patient's cells, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR prior to analysis (Saiki et al., *Nature*, 324: 163-166 (1986)). Ligation-mediated amplification may also be used for amplification (Vollach, V., et al., *Nucl. Acids Res.* 22: 2507 (1994)). RNA or cDNA may also be used in the same ways. As an example, PCR

primers complementary to the nucleic acid encoding cathepsin K can be used to identify and analyze cathepsin K expression and mutations. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled cathepsin K RNA or alternatively, radiolabeled cathepsin K antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

Sequence differences between a reference gene and genes having mutations also may be revealed by direct DNA sequencing. In addition, cloned DNA segments may be employed as probes to detect specific DNA segments. The sensitivity of such methods can be greatly enhanced by appropriate use of PCR or another amplification method. For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotide or by automatic sequencing procedures with fluorescent-tags.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, e.g., Myers et al., Science, 230: 1242 (1985)).

Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (e.g., Cotton et al., Proc. Natl. Acad. Sci., USA, 85: 4397-4401 (1985)).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (e.g., restriction fragment length polymorphisms ("RFLP"), SSCP and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations also can be detected by in situ analysis.

Chromosome assays

The sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

In certain preferred embodiments in this regard, the gDNA herein disclosed is used to clone genomic DNA of a cathepsin K gene. This can be accomplished using a variety of well known techniques and libraries, which generally are available commercially. The genomic DNA then is used for in situ chromosome mapping using well known techniques for this purpose. Typically, in accordance with routine procedures for chromosome mapping, some trial and error may be necessary to identify a genomic probe that gives a good in situ hybridization signal.

In some cases, in addition, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the gDNA. Computer analysis of the 3' untranslated region of the gene is used to rapidly select primers that do not span more than one exon in the genomic DNA complicate the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes (e.g., radiation hybrid panels) or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include in situ hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence in situ hybridization ("FISH") of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step.

This technique can be used with gDNA as short as 50 to as long as 600. For a review of this technique, see Verma et al., HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES, Pergamon Press, New York (1988).

5 Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, MENDELIAN INHERITANCE IN MAN, available on line through Johns Hopkins University, Welch Medical Library. The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent
10 genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

15 With current resolution of physical mapping and genetic mapping techniques, a gDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb).

20 Polypeptide assays

The present invention also relates to a diagnostic assays such as quantitative and diagnostic assays for detecting levels of cathepsin K protein in cells, tissues and bodily fluids, including determination of normal and abnormal levels of polypeptide. Bodily fluids useful in the diagnostic methods of the invention include, for example, synovial fluid,
25 cerebrospinal fluid, urine, serum, gingival fluid and lymph. Thus, for instance, a diagnostic assay in accordance with the invention for detecting over-expression of cathepsin K protein compared to normal control tissue samples may be used to detect the presence of disease, for example, osteoporosis, Paget's disease, Gaucher's disease, CNS inflammation, Alzheimer's disease, hyperparathyroidism, bone degradation, metastatic tumors, rheumatoid
30 arthritis, osteoarthritis, periodontal disease and degradation of bone implants and bone protheses, particularly dental implants. Assay techniques that can be used to determine levels of a protein, such as an immunoassay for cathepsin K protein of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot

analysis and ELISA assays. Among these ELISAs frequently are preferred. An ELISA assay initially comprises preparing an antibody specific to cathepsin K, preferably a monoclonal antibody. In addition a reporter antibody generally is prepared which binds to the monoclonal antibody. The reporter antibody is attached a detectable reagent such as radioactive, fluorescent or enzymatic reagent, in this example horseradish peroxidase enzyme.

To carry out an ELISA a sample is removed from a host and incubated on a solid support, e.g. a polystyrene dish, that binds the proteins in the sample. Any free protein binding sites on the dish are then covered by incubating with a non-specific protein such as bovine serum albumin. Next, the monoclonal antibody is incubated in the dish during which time the monoclonal antibodies attach to any cathepsin K proteins attached to the polystyrene dish. Unbound monoclonal antibody is washed out with buffer. The reporter antibody linked to horseradish peroxidase is placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to cathepsin K. Unattached reporter antibody is then washed out. Reagents for peroxidase activity, including a colorimetric substrate are then added to the dish. Immobilized peroxidase, linked to cathepsin K through the primary and secondary antibodies, produces a colored reaction product. The amount of color developed in a given time period indicates the amount of cathepsin K protein present in the sample. Quantitative results typically are obtained by reference to a standard curve.

A competition assay may be employed wherein antibodies specific to cathepsin K attached to a solid support and labeled cathepsin K and a sample derived from the host are passed over the solid support and the amount of label detected attached to the solid support can be correlated to a quantity of cathepsin K in the sample.

25

Antibodies

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of a Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or

by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue
5 expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous clonal cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature 256: 495-497 (1975), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., Immunology Today 4: 72
10 (1983) and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., pg. 77-96 in MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc. (1985).

Techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide
15 products of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express antibodies, including for example, humanized antibodies to immunogenic polypeptide products of this invention.

Thus, among others, such antibodies can be used to detect and treat diseases caused by or associated with mutant cathepsin K or abnormal cathepsin K levels, such as,
20 osteoporosis, periodontal disease, Paget's disease, Gaucher's disease, CNS inflammation, Alzheimer's disease, hyperparathyroidism, bone degradation, metastatic tumors, and degradation of bone implants and bone prostheses, particularly dental implants.

Immunization using polynucleotides of the inventions can be carried out using known methods to produce a cathepsin K-specific immune response.
25

Clinical Genomics

This invention provides methods to determine drug responsiveness of individuals having or suspected of having a cathepsin K gene mutation or cathepsin K gene expression abnormality, and also provides reagents to carry out such methods. Individuals may be
30 grouped by their responsiveness to a given compound, particularly drugs, used to treat diseases caused by or associated with a mutation of cathepsin K gene or cathepsin K gene expression. Such individuals may be further grouped by detecting different gene mutations or gene expression level variants. In this manner specific gene mutations and gene expression variants can be readily associated with a certain degree of responsiveness to a

compound by an individual). Methods and reagents provided herein can be used to group compound responsiveness by detecting cathepsin K gene mutations and cathepsin K gene expression variants. Other methods for grouping individuals by compound responsiveness are known to skilled artisans and can be adapted to use the polypeptides and polynucleotides of the invention.

The invention also provides algorithms useful in conjunction with a device or embodied in a composition matter which are useful for the diagnosis of diseases caused by or associated with cathepsin K or mutants or variants thereof. Preferred algorithms are provided for disease stratification and staging.

Cathepsin K binding molecules and assays

This invention also provides a method for identification of molecules, such as receptor molecules, that bind cathepsin K or fragments of cathepsin K of the invention. Genes encoding proteins that bind cathepsin K, such as receptor proteins, can be identified by numerous methods known to those of skill in the art, for example, ligand panning and FACS sorting. Such methods are described in many laboratory manuals such as, for instance, Coligan et al., Current Protocols in Immunology 1(2): Chapter 5 (1991).

For instance, expression cloning may be employed for this purpose. To this end polyadenylated RNA is prepared from a cell responsive to cathepsin K, a cDNA library is created from this RNA, the library is divided into pools and the pools are transfected individually into cells that are not responsive to cathepsin K. The transfected cells then are exposed to labeled cathepsin K. (Cathepsin K can be labeled by a variety of well-known techniques including standard methods of radio-iodination or inclusion of a recognition site for a site-specific protein kinase.) Following exposure, the cells are fixed and binding of cathepsin K, or a molecule which binds to cathepsin K, is determined. These procedures conveniently are carried out on glass slides.

Pools are identified of cDNA that produced cathepsin K-binding cells. Sub-pools are prepared from these positives, transfected into host cells and screened as described above. Using an iterative sub-pooling and re-screening process, one or more single clones that encode the putative binding molecule or substrate, such as cell matrix, bone matrix or a receptor molecule, and the any of the same can be isolated.

Alternatively a labeled ligand can be photoaffinity linked to a cell extract, such as a membrane or a membrane extract, prepared from cells that express a molecule that it binds, such as a receptor molecule. Cross-linked material is resolved by polyacrylamide gel

electrophoresis ("PAGE") and exposed to X-ray film. The labeled complex containing the ligand-receptor can be excised, resolved into peptide fragments, and subjected to protein microsequencing. The amino acid sequence obtained from microsequencing can be used to design unique or degenerate oligonucleotide probes to screen cDNA libraries to identify
5 genes encoding the putative receptor molecule.

Polypeptides of the invention also can be used to assess cathepsin K binding capacity of cathepsin K binding molecules, such as receptor molecules, in cells or in cell-free preparations.

10 **Agonists and antagonists - assays and molecules**

The invention also provides a method of screening compounds to identify those which enhance or block the action of cathepsin K in or on cells, such as its interaction with cathepsin K-binding molecules such as receptor and enzymatic substrate molecules. An agonist is a compound which increases the natural biological functions of cathepsin K,
15 while antagonists decrease or eliminate such functions.

For example, a cellular compartment, such as a membrane, vacuole, inclusion or a preparation of any thereof, such as a membrane-preparation, may be prepared from a cell that expresses a molecule that binds cathepsin K, such as a molecule of a signaling or regulatory pathway modulated by cathepsin K. The preparation is incubated with labeled
20 cathepsin K in the absence or the presence of a candidate molecule which may be a cathepsin K agonist or antagonist. The ability of the candidate molecule to bind the binding molecule, such as a substrate, is reflected in decreased binding of the labeled ligand. Molecules which bind gratuitously, i.e., without inducing the effects of cathepsin K on binding the cathepsin K binding molecule, are most likely to be good antagonists.
25 Molecules that bind well and elicit effects that are the same as or closely related to cathepsin K are agonists.

Cathepsin K-like effects of potential agonists and antagonists may be measured, for instance, by determining activity of a second messenger system following interaction of the candidate molecule with a cell or appropriate cell preparation, and comparing the effect
30 with that of cathepsin K or molecules that elicit the same effects as cathepsin K. Second messenger systems that may be useful in this regard include but are not limited to AMP guanylate cyclase, ion channel, phosphoinositide hydrolysis second messenger systems, or compounds which signal the binding of a potential agonists and antagonists to cathepsin K or its substrate.

Another example of an assay for cathepsin K antagonists is a competitive assay that combines cathepsin K and a potential antagonist with enzymatic substrate or substrate analogs under appropriate conditions for a competitive inhibition assay. Cathepsin K can be labeled, such as by radioactivity, such that the number of cathepsin K molecules bound to a receptor molecule can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a receptor molecule, without inducing cathepsin K-induced activities, thereby preventing the action of cathepsin K by excluding cathepsin K from binding.

Potential antagonists include a small molecule which binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such as receptor molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules.

Other potential antagonists include antisense molecules. Antisense technology can be used to control gene expression through antisense DNA or RNA or through triple-helix formation. Antisense techniques are discussed, for example, in - Okano, J. Neurochem. 56: 560 (1991); OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988). Triple helix formation is discussed in, for instance Lee et al., Nucleic Acids Research 6: 3073 (1979); Cooney et al., Science 241: 456 (1988); and Dervan et al., Science 251: 1360 (1991). The methods are based on binding of a polynucleotide to a complementary DNA or RNA. For example, the 5' coding portion of a polynucleotide that encodes the mature polypeptide of the present invention may be used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription thereby preventing transcription and the production of cathepsin K. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of the mRNA molecule into cathepsin K polypeptide. The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of cathepsin K.

The antagonists may be employed in a composition with a pharmaceutically acceptable carrier, e.g., as hereinafter described.

The antagonists may be employed for instance to treat diseases caused by or associated with mutant cathepsin K or abnormal cathepsin K levels, such as, osteoporosis, 5 Paget's disease, Gaucher's disease, CNS inflammation, Alzheimer's disease, hyperparathyroidism, bone degradation, metastatic tumors, rheumatoid arthritis, osteoarthritis, periodontal disease and degradation of bone implants and bone prostheses, particularly dental implants.

10 Compositions

The invention also relates to compositions comprising the polynucleotides or the polypeptides discussed above or the agonists or antagonists. Thus, the polypeptides of the present invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for 15 administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration.

20

Kits

The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Associated with such container(s) can be a notice in the form 25 prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, reflecting approval by the agency of the manufacture, use or sale of the product for human administration.

Administration

30 Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous,

intraperitoneal, intramuscular, subcutaneous, intranasal, intraarticular, or intradermal routes among others.

The pharmaceutical compositions generally are administered in an amount effective for treatment or prophylaxis of a specific indication or indications. In general, the compositions are administered in an amount of at least about 10 mg/kg body weight. Preferably, in most cases, dose is from about 10 mg/kg to about 1 mg/kg body weight, daily. It will be appreciated that optimum dosage will be determined by standard methods for each treatment modality and indication, taking into account the indication, its severity, route of administration, complicating conditions and the like.

10

Gene therapy

The cathepsin K polynucleotides, polypeptides, agonists and antagonists that are polypeptides may be employed in accordance with the present invention by expression of such polypeptides *in vivo*, in treatment modalities often referred to as "gene therapy."

15

Thus, for example, cells from a patient may be engineered with a polynucleotide, such as a DNA or RNA, encoding a polypeptide *ex vivo*, and the engineered cells then can be provided to a patient to be treated with the polypeptide. For example, cells may be engineered *ex vivo* by the use of a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention. Such methods are well-known in the art and their use in the present invention will be apparent from the teachings herein.

20

Cells from a patient may also be engineered with a polynucleotide, such as a ribozyme that has been constructed, using well known methods, to inhibit the gene expression of Cathepsin K. Other constructs may also be engineered into a patient's cells to contain antisense stretches of cathepsin K sequence, using well known methods. Such antisense constructs will inhibit Cathepsin K expression in the patient.

25

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by procedures known in the art. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct then may be isolated and introduced into a packaging cell that is transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and

30

other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention.

Retroviruses from which the retroviral plasmid vectors herein above mentioned may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen
5 necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus. In one embodiment, the retroviral plasmid vector is derived from Moloney Murine Leukemia Virus.

Such vectors will include one or more promoters for expressing the polypeptide.
10 Suitable promoters which may be employed include, but are not limited to, cathepsin K promoter, a retroviral LTR, an SV40 promoter, and the human cytomegalovirus (CMV) promoter described in Miller et al., Biotechniques 7: 980-990 (1989), or any other promoter (e.g., cellular promoters such as eukaryotic cellular promoters including, but not limited to, the histone, RNA polymerase III, and alpha-actin promoters). Other viral promoters which
15 may be employed include, but are not limited to, adenovirus promoters, thymidine kinase (TK) promoters, and B19 parvovirus promoters. The selection of a suitable promoter will be apparent to those skilled in the art from the teachings contained herein.

The nucleic acid sequence encoding the polypeptide of the present invention will be placed under the control of a suitable promoter. Suitable promoters which may be
20 employed include, but are not limited to, adenoviral promoters, such as the adenoviral major late promoter; or heterologous promoters, such as the cytomegalovirus (CMV) promoter; the rous sarcoma virus (RSV) promoter; inducible promoters, such as the MMT promoter, the metallothionein promoter; heat shock promoters; the albumin promoter; the ApoA1 promoter; human globin promoters; viral thymidine kinase promoters, such as the
25 Herpes Simplex thymidine kinase promoter; retroviral LTRs (including the modified retroviral LTRs herein above described); the alpha-actin promoter; and human growth hormone promoters. The promoter also may be the native promoter which controls the gene encoding the polypeptide.

The retroviral plasmid vector is employed to transduce packaging cell lines to form
30 producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the PE501, PA317, Y-2, Y-AM, PA12, T19-14X, VT-19-17-H2, YCRE, YCRIP, GP+E-86, GP+envAm12, and DAN cell lines as described in Miller, A., Human Gene Therapy 1: 5-14 (1990). The vector may be transduced into the packaging cells through any means known in the art. Such means

include, but are not limited to, electroporation, the use of liposomes, and CaPO_4 precipitates. In one alternative, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

- The producer cell line will generate infectious retroviral vector particles, which
- 5 include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed to transduce eukaryotic cells, either *in vitro* or *in vivo*. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells,
- 10 hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

EXAMPLES

- 15 The present invention is further described by the following examples. The examples are provided solely to illustrate the invention by reference to specific embodiments. These exemplifications, while illustrating certain specific aspects of the invention, do not portray the limitations or circumscribe the scope of the disclosed invention.

- 20 Certain terms used herein are explained in the foregoing glossary.

An N used herein in a nucleotide sequence refers to an unknown nucleotide or nucleotides.

All examples were or may be carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail.

- 25 Routine molecular biology techniques of the following examples can be carried out as described in standard laboratory manuals, such as Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), herein referred to as "Sambrook."

- All parts or amounts set out in the following examples are by weight, unless
- 30 otherwise specified.

Unless otherwise stated size separation of fragments in the examples below was carried out using standard techniques of agarose and polyacrylamide gel electrophoresis ("PAGE") in Sambrook and numerous other references such as, for instance, by Goeddel et al., *Nucleic Acids Res.* 8: 4057 (1980).

Unless described otherwise, ligations were accomplished using standard buffers, incubation temperatures and times, approximately equimolar amounts of the DNA fragments to be ligated and approximately 10 units of T4 DNA ligase ("ligase") per 0.5 µg of DNA.

5

Example 1 Isolation and sequencing of human cathepsin K genomic clone

- cDNA as disclosed in U.S. Patent Number 5,501,969, was used to isolate the gDNA clone from a gDNA library (Clontech) according to the following method. Primers to adjacent exons (6 of the 7 exons) were prepared. The sequence of these primers is underlined in Figure 2. PCR was performed using standard methods well known in the art. Amplified fragments were cloned into a TA vector (Clontech) and the clones were sequenced by an automated sequencer (Applied BioSystems Model 373) by established methods well known in the art using forward and reverse sequencing primers. The
- 10 sequence of all internal introns were obtained. 5' and 3' terminal intron sequences were obtained as follows. 5' end primers were designed to obtain sequence for the first intron (see underlined primer in Figure 2), using these primers 2 P1 clones were obtained (Genome Systems Inc.). Both clones were full length. PCR was used to confirm the sequence of internal intron-exon boundary junctions (see Example 2). Primers derived from
- 20 sequence at the 5' end of the P1 clones was used to "walk" and sequence along the clone, in a stepwise fashion, using new primers at each sequence step, by routine methods known in the art. Purification of P1 clones was carried out as set forth in Example 1(d). "Walking" and sequencing was performed in both directions to confirm cathepsin K gDNA sequence. PCR was again performed using proofreading Taq polymerase (PCR Ultima, Perkin Elmer).
- 25 A transcription start site was obtained using a 5' RACE kit (Gibco BRL) and the protocol supplied therewith. This site was also confirmed using an RNase protection assay kit (Hybspeed, RPA Ambion). Example 1 (a)-(d) provide further specifics concerning cloning and sequencing of cathepsin K

(a) DNA sequencing of Intron-exon boundaries

Intron-Exon Boundaries

5 **Intron 1**

Intron 1 was identified by utilization of 5' RACE (Gibco BRL) technique to determine 5' UTR sequence from which primer could be designed to PCR from exon 1 to exon 2. (intron 1 starts prior to ATG so PCR may not be readily employed based on cDNA sequence available.) Intron 1 was amplified by PCR on human genomic DNA (Clontech) and cloned into PCRII vector and sequenced as described in Example 1.

Intron 2

Intron 2 was identified by PCR on human genomic DNA from primers designed in exon 2 to exon 3. PCR product was cloned and sequenced using standard methods.

Intron 3

Intron 3 was identified by PCR on human genomic DNA from primers designed in exon 3 to exon 4. PCR product was cloned and sequenced using standard methods.

20 **Intron 4**

Intron 4 was identified by PCR on human genomic DNA from primers designed in exon 4 to exon 5. PCR product was cloned and sequenced using standard methods.

Introns 5 & 6

25 Introns 5 and 6 were identified by PCR on human genomic DNA from primers designed in exon 5 to exon 7. PCR product was cloned and sequenced using standard methods confirming presence of both introns.

Intron 7

30 Intron 7 was identified by PCR on human genomic DNA from primers designed in exon 7 to exon 8. PCR product was cloned and sequenced using standard methods.

All introns that were identified by PCR on human genomic DNA were confirmed by PCR of the same regions on P1 clone A (see (b) below) clone (Genome Systems, Inc.)

(b) DNA sequencing of 5' and 3' untranslated region (UTR)

5' and 3' untranslated regions were isolated from a single P1 clone (Genome Systems Inc.). This P1 clone has been identified herein as "P1 clone A." Sequence was
5 obtained by direct sequence walking up and down the P1 clone with gene specific primers derived from confirmed cDNA sequence using standard methods. These regions were then cloned via PCR and confirmed by sequence analysis using standard methods. The 5' UTR was additionally amplified by PCR using proofreading Taq Polymerase Ultima in accordance with manufacturer instructions and cloned to eliminate sequence ambiguities. 5'
10 and 3' UTR were further confirmed by PCR on human genomic DNA using standard methods.

(c) DNA sequencing of mRNA Cap Site & size of exon 1

An mRNA Cap Site was determined to be about 48 bp upstream of the start codon
15 based on 5' RACE sequencing. Ribonuclease Protection Assay confirmed a protected fragment of about 48 bp in size indicating that the start site from transcription resides about 48 bp upstream of ATG (start codon). Putative transcription factors have been identified by analysis of sequence with database transcription factor sequence information and these are set forth in Figure 3(S) [SEQUENCE ID NO: 2]. The 1.1 kb 5' UTR fragment was cloned
20 into pCAT expression vectors to further analyze the promoter sequence region.

(d) P1 DNA preparation

P1 clone A colonies were streaked out on Kanamycin LB plates. A single colony was picked and grown O/N in 20 mls with 25 µg/ml of kanamycin. 500 mls of media (25
25 µg/ml kanamycin) was inoculated with 16 mls of the O/N culture and grown for 10 hours. Cells were pelleted by centrifugation and resuspended in 10 mls of Qiagen P1 Solution. 10 mls of Qiagen P2 Solution was added and incubated at room temp. for 5 min. 10 mls of Qiagen P3 Solution was added and the mixture left on ice for 15 min. The sample was spun at 10,000g for 15 min. The supernatant was removed and extracted with phenol. The
30 supernatant was then re-extracted with chloroform. The DNA was precipitated following addition of NaOAc pH 5.2 and 1.1 volumes of isopropanol. The DNA was pelleted by centrifugation for 15 min. at 10,000g and washed with 70% ethanol. To clean up the DNA for sequencing, 250 µl of DNA (about 50 µg) was added to 65 µl 30% PEG in 1.5M NaCl. 8.5 µl of 3M NaCl was added and the mixture incubated on ice for 30 min. The sample was

spun at 12,000g for 10 min. The supernatant was discarded and the pellet dissolved in 200 μ l distilled water. The DNA was then extracted with chloroform, vortexed and spun at 10,000g for 1 min. The aqueous layer was removed and the DNA precipitated with 40 μ l of NaOAc pH 5.2 and 1 ml of ethanol. The sample was spun at 12,000g for 30 min. The DNA was washed with 1 ml of 70% ethanol and resuspended on distilled water. Prior to sequencing, the DNA was denatured with 0.1 volumes of 2M NaOH and 2mM EDTA and incubated at 37°C for 30 min. The mixture was neutralized with 0.1 volume of 3M NaOAc pH 5.2 and precipitated with 2.5 volumes of ethanol. The denatured DNA was resuspended in distilled water at a concentration of 1 μ g/ μ l. 6 μ g/ μ l were used in each sequencing reaction (ABI) using TaqFS.

Example 2 Chromosomal mapping of cathepsin K

Purified P1 DNA was used for FISH analysis (Genome Systems, Inc.) to map to specific chromosome. Prior results done showed by use of 2 PCR somatic cell hybrid panels that the gene mapped to Chromosome 1. FISH analysis confirmed mapping to chromosome 1 and also further mapped the gene to 1q21. This is the same locus as is known for cathepsin-S.

Example 3 Expression of cathepsin K in COS cells

CAT Assays

pCAT-CatK, which contains the 1100 bp putative CatK promoter, upstream of the CAT reporter gene was transfected into COS cells by the DEAE-dextran procedure. Transfections were done on COS cells in 100mm dishes and 5 μ g of DNA was used. As controls, pCAT basic, which contains no promoter or enhancer, and pCAT control, which contains the SV40 promoter and enhancer, were also transferred separately. 72 hours after transfection, extracts were made by freeze-thaw and equal amounts of extract protein were used in both 1-hour and overnight CAT assays. No activity was detected in untransfected COS cells. pCAT-CatK showed a 1.4-1.6 fold increase of CAT expression relative to pCAT basic after subtraction of background from untransfected cells. Since it is possible that higher levels of activation may be obtained in the presence of various inducers, activation of the CatK promoter by adding exogenous 1,25 di-hydroxy vitamin D3 is believed to occur. Vitamin D has been shown by others to activate transcription of osteocalcin, osteopontin, calcitonin and P450 promoters through interaction with the vitamin D receptor and the vitamin D

response element(s) found in these various promoters. The ability of vitamin D to transactivate these promoters is believed thought to play a role in the control of bone formation and resorption. Similar experiments can be performed to assess estrogen responsiveness which is also believed thought to play a role in the control of bone formation and resorption.

Example 4 Gene therapeutic expression of human cathepsin K

Fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature overnight. After 24 hours at room temperature, the flask is inverted - the chunks of tissue remain fixed to the bottom of the flask - and fresh media is added (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin). The tissue is then incubated at 37°C for approximately one week. At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerges. The monolayer is trypsinized and scaled into larger flasks.

A vector for gene therapy is digested with restriction enzymes for cloning a fragment to be expressed. The digested vector is treated with calf intestinal phosphatase to prevent self-ligation. The dephosphorylated, linear vector is fractionated on an agarose gel and purified.

Cathepsin K gDNA capable of expressing active cathepsin K, is isolated. Preferred constructs use the cathepsin K promoter for cell type-specific gene expression. The ends of the fragment are modified, if necessary, for cloning into the vector. For instance, 5' overhanging may be treated with DNA polymerase to create blunt ends. 3' overhanging ends may be removed using S1 nuclease. Linkers may be ligated to blunt ends with T4 DNA ligase.

Equal quantities of the Moloney murine leukemia virus linear backbone and the cathepsin K fragment are mixed together and joined using T4 DNA ligase. The ligation mixture is used to transform *E. Coll* and the bacteria are then plated on agar-containing kanamycin. Kanamycin phenotype and restriction analysis confirm that the vector has the properly inserted gene.

Packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin.

The vector containing the cathepsin K gene is introduced into the packaging cells by standard techniques. Infectious viral particles containing the cathepsin K gene are collected from the packaging cells, which now are called producer cells.

5 Fresh media is added to the producer cells, and after an appropriate incubation period media is harvested from the plates of confluent producer cells. The media, containing the infectious viral particles, is filtered through a Millipore filter to remove detached producer cells. The filtered media then is used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the filtered media. Polybrene (Aldrich) may be included in the media to facilitate transduction. After
10 appropriate incubation, the media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his, to select out transduced cells for expansion.

Engineered fibroblasts then may be injected into humans or animals, including for
15 example, rats and mice, either alone or after having been grown to confluence on microcarrier beads, such as cytodex 3 beads. The injected fibroblasts produce cathepsin K product, and the biological actions of the protein are conveyed to the host.

It will be clear that the invention may be practiced otherwise than as particularly
20 described in the foregoing description and examples.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Debouck, Christine
Drake, Fred
Gowen, Maxine
Rood, Julie
Hastings, Gregg
Adams, Mark
Fraser, Claire
Lee, Norman
Kirkness, Ewen
Blake, Judith
Fitzgerald, Lisa

(ii) TITLE OF THE INVENTION: CATHEPSIN K GENE

(iii) NUMBER OF SEQUENCES: 32

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19406-2799

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5.

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE: 07-MAY-1997
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 60/019,942
(B) FILING DATE: 14-JUNE-1996

整理番号 156618

PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 60/020,273
(B) FILING DATE: 17-JUNE-1996

PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 60/026,273
(B) FILING DATE: 26-AUG-1996

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Han, William T
(B) REGISTRATION NUMBER: 34,344
(C) REFERENCE/DOCKET NUMBER: ATG50006-2

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 610-270-5219
(B) TELEFAX: 610-270-5090
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13674 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(x) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCTTTGGCTC CCAAAGGCCT GGGATTACAG GCGTGAACCA CTGCCCTAG CCTGTTAGCA	60
GCTCTTAAAA TCCAGAGGCA TAAGCCTGTA TTTTGGAGGG TTTATGCATG GAATCCAGCT	120
AGAACTGAG TCTATTACAG ATCCCATTTA TTATCCTTTC TATTCCAAGA AGCCTTTTTT	180
TCTCCTCCC CACATCTGTT TATGGAAGAA AATGAAGTTT GGGGTGTGGT TTGAGGAATC	240
AGCTAGATTC TTATGATCTG TCACATGCTT GGATGTTGGG GAAGCATTTG GAGAAGCTCA	300
TGTGACTTGT CCTAGATTGG GGATTTTAAT TGAGACAGAT GATGTTTATC GGGCATCCCA	360
CCACCTGAGA GTTTTAGCAA CAGAGTCACA TGTGAGTCCA TCAGAACTTA CGGCATTGAT	420
TCAAGTGCTG TCATAAATAA CCAGGACTGC TGTTTTGGT TACTTTTAAA GACAGTTTCA	480
TCTGGACTTT CTGGGCATAT CCTCCTCAG CAAAACCACA TTAGGCTGGG AAAACTATTC	540

整理番号 156518

TGCTTGGGAG TAATGACAAC TTGCAACCAA CAGCTTATA AAAATACAAA GAATTCTGGA	600
GCCTATGGCT TCCATTACAT TATTCTTTTA TAGCCTTTTA TGTTTCATTAC CGCATCCCAG	660
AGCTGAGAGT CAGACACAAA TATGAAAATA GGTTCATATG TTGGAGAGGT AAATCCTAAC	720
AGGAAAGGGG TAGGAAAAGA TATATATCCC CAATATTAAA ATAAAGATAT TGAAGAAGAA	780
GGATGGGAGA GACTAGGGCT GTGTCTTCC TTTACTCAC CAAAAGAGAA AGTAAGCTCC	840
TATTTGAGTC AATAGATATT GAGGTCTTGT TATTTGCCAC CAAAGACAGT CTTGTGAGAC	900
TAAATAGCTA GTAATTCCCT ACCCTGGCAC ACATGCTGCA TACACACAGA AACACTGCA	960
ATCCACTGCC TCCCTCCCTC CTCCTACCC TCCCTCTCT CAGCATTTCT ATCCCCGCT	1020
CCTCCTCTTA CCCAAATTTT CCAGCCGATC ACTGGAGCTG ACTTCCGCA TCCCGATGGA	1080
ATAAATCTAG CACCCTGAT GGTGTGCCA CACTTTGCTG CCGAAACGAA GCCAGACAA	1140
AGATTTCAT CAGCAGGTAA CGTTTGCAAC TCCTAGATC TTTTAGCTTT TCATTCTGT	1200
CAATTCTCTG AGTATTAGGG ATGTAGTGAC TTGAGGATCA CAATAAACTT TTAGCCTCTG	1260
CAGATGAAAA CAGAGATGCA CTTCTTAGGT CATTCCCTGG CTAAATAAAA TCTGCCGGA	1320
AATCTGTAGA ATTCCTTGTA TGATTTATAT ATATACATAC ATGATTGTGA GTAAAAGCAA	1380
AGTATATAGG GAATCATTTC CCCATCCTTC AAGAGTGGCC TTTCTGCAGT GTTTTCTACT	1440
TTGGCCAACA AGGATCAAAA CGGTAACTC CTTAGTGAGG AGGAGGAGAG TGGTATGGG	1500
AGGTAGTAGC TCAGTGCTTC CTGTTCAGTG AGACATCTCA AAGCCCTTAA CACTCTAGTT	1560
TTTAAATGTC CTACTGGACA TTTTGCCAGT TTGCAAAAT ACATGTAAAT GGACTATAAG	1620
CAATTGTGTA AGCCATATGT CATGTGCAG GCTGCAAAAT GTTCTTAAA TGGAGGATTT	1680
GTAATTAAGA AAGCCAATGC AAGAAATGAG TGAAGCTAAC TAGACTAAC TTATGAAAAG	1740
CTGTGAATTT CATCATATA GAACATTGCT TTTGAGTCTG AACATTCTTC TAACAAACCT	1800
TGGATCTGAG GCTTCTTGTC CTTTGGGCA GCCACAGTGG GTTTTGTG TTAGGGGAAA	1860
ATAAAAAACC TTGCCCGCAG CATCTGGTTA AGATTAGGGC AGTTTCTGC CTAAGGAGG	1920
AAGGGAGAGA AAAAGGAAGA AGAAATGCAT AAGGAGAAAT AGGAGATATA CAATCTCTCA	1980
GAAAACAGGA AACATTGTCC TATTTTCCCT TGTCTCTTC TGACAAGATC TGGGAAAGTA	2040
CCAGAAATTA GGCACGAAAG AGAAGAACGC CTCGAAGAAA TGATCAGGAA GCAAACTTA	2100
GACGGAATC TCTCCTTGT GTATTCGAA CCCCACTACC ACCTTGCTAT TTGTCTGTCT	2160
CCAAGCCTGC TAGGGACCTT GGAGGAAACG CACTGAGCCC ATCTGATTG TCCAGTTTCT	2220
ATCCCCCATT TCTGGTTGTG TACGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG	2280
TGTGTGTGAG AGAGAGAGAG ACAGAGAGAG AACAGAGAG AGTGTGTGTT GCCTAAATCT	2340
CCCAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAAAAGAGA GAAATGGCTA	2400
AATCCCCCTA GATCAAAGTC CTTGGAACCA GATGTACCAG CATCTATCT AAACACAGGC	2460
CCCTCCTGAC TATCATTGTT TTATCACCTT TTTCCGTCT ACCTTCTCT TCTCATAAA	2520
GCCTAGTTTT CCTCTGTTTC CTGCCAAAT GGAAGAGTTT TCCCTAATA CATTCTCTG	2580
CAGGATGTGG GGGCTCAAGG TTCTGCTGCT ACCTGTGGTG AGCTTTGCTC TGTACCCTGA	2640
GGAGATACTG GACACCCACT GGGAGCTATG GAAGAAGACC CACAGGAAGC AATATAACAA	2700
CAAGGTGCCT GGGGTCTGG AGGGGGCATG GCAGGAAGGC TGAGACCTGA GCTCTCTCAT	2760
CTTAGCTTCC AGACTCCCTT CTCAATCCA AATGCTTTAT TCCAAGCAA TCACTCCCTC	2820
TTCCCTAACT CATGTTAACA TACGGTTTTC ATTCCTATGC TTCAATCATC CTCTTGTCAA	2880
ACTTGTATTC CTTCCCTTTG GTTTTATAAG TGTGTAACAT TCCTCTTTTG GGAAGAGTCC	2940
CAAGATTAAAT GCTGTTAATC CATAAGCAAT TTTTCTGTCT CTCCAGAGCT TGTGTGTTG	3000
TTTACATATT ATCTCTCTTC TTGCAGGCTC TTAATTCAT GGTAGTTCC CCAACTAAAC	3060
TGTAAACTTT TATGATTGTG AGTTTCCTTT ATTCTCTTAA AACCTTCAC AATATTACAT	3120
ATGAACGTGA GACAGTCTAT ACAAGTACTG ACTATGCTTT GTTTAGGTGG ATGAAATCTC	3180

整理符号 156518

TCGGCGTTTA	ATTGCGGAAA	AAAACCTGAA	GTATATTTC	ATCCATAACC	TTGAGGCTTC	3240
TCTTGCTGTC	CATACATATG	AACTGGCTAT	GAACCACCTG	GGGACATGG	CAAGTATAGC	3300
TTCAGCTCCT	GTCCACCTG	CACCATTTGC	TTAGTTCCC	TGCTGATGCC	TGGCCTCTTT	3360
CTTCTTTGTC	TTAGACCAGT	GAAGAGGTGG	TTCAGAAGAT	GACTGGACTC	AAAGTACCCC	3420
TGCTCATTC	CCGCACTAAT	GACACCTTT	ATATCCCAGA	ATGGGAAGGT	AGAGCCCCAG	3480
ACTCTGTCGA	CTATCGAAAG	AAAGGATATG	TTACTCCTGT	CAAAAATCAG	GTACTCTCCT	3540
TTCTTCTGGG	TGTGCATATG	TAATCTGGCA	TGACCTTTTC	CTTTTCTGTC	TGCTTTGTTT	3600
TTGAGGTGAA	AGGGCACCAG	GAAAAGAGCG	CAAGGAATTA	AGGTACATCT	CCCCATTCCT	3660
ATTCTGTTAT	TTAACCTCAT	TTGTTCTGT	ACATTTGGGT	TGTTTCTGGT	TTTTCTTTTT	3720
CTTTTCCCTT	TTTTTTTTTT	TTTTTTTTTT	GAGATAGAGT	CTCACTCTGT	CGCCAGGAT	3780
GGAGTGCACT	GGTGCAATCT	TGGCTCACTG	CAACCTACAC	CTCCCGGTT	CAAGCGATTC	3840
TCCTGCCTCA	GCCTCCTGAG	TAGCTGAGAT	TACAGGCACG	CGCCACTACG	CCTGGCTAAT	3900
TTTTCTATTT	TTATAGAGAT	GCGTTTTCAC	CATGTTGGCC	AGGCTGGTCT	TGAACGTACC	3960
TCAGGTGATC	CACCTGCCTC	AGCCTCCCAA	AGTGCTGGGA	TTAGAGTCAT	GAGCCATCGC	4020
GGCCTGGTTT	TTCTTTATTA	CAATAGTGT	TGCAATAAGC	ACCCTTGTGC	ATATGTTTTT	4080
GTGCACATGT	ACAAATATTT	ATGCAAAATA	AGTCCTAAAA	TTGGAATTGT	TAGGTCACAA	4140
ATAATCCTTT	CCCCCCCCC	AAATTTTTTT	TTTTTTTTTT	AGACAGCGTC	TCTGTCACCC	4200
AGGCTGGAGT	CCAGTGGCGC	AATCATGGCT	CACCTGCACG	TCAACGTCTC	AGGCTCAAGT	4260
GATTTCTCAA	CCTCAGCCTC	CCTAGTAGCT	GGAATTTAGA	AGCACATGCC	ACCACACCCA	4320
GCTAATTTTA	AAAAATTTTT	TGTTAGAGAC	AGGCTTTTGC	CATGCTACCC	AAGCTGGTCT	4380
CAAATTCCTG	GGCTCAAGCA	ATCTGCCCGC	TTCCGCCCTCC	CAAAGTGCTA	GGATTACAGA	4440
CATGAGCCAC	CATGCCACGC	CCAAAAAGT	TTTTGCAATC	TTACATTCTT	ACTAGCATGA	4500
GATGTCAGT	TTTTTCACAA	CCCAACAAC	ACAGGATTGT	ATCAGCAAGA	TAAACAATTG	4560
ATTTAACGTT	CATTTAACAA	ACACTTTTTG	ACCCCCAGAA	CCTACCAGAT	GCAGTGTAG	4620
GCAGCAGAGA	CTCAAGATGA	CTAAGACACA	ACCTGTGTCC	TCAGGAATTC	TCAATCTAAA	4680
AAAATAGAAC	AGGAAAGAAA	GAAAAATCTA	CAATCTAGCT	GCACAAACAA	TAATAGCTAA	4740
TACTTTTTGA	GATTTIATTT	TTTGTCAGGA	ACTTCTTAAC	TCTTTACATG	AGTTTAAATA	4800
TTTAATCCCT	TATAACAATA	TTTTATGCAT	AGAGAAACTG	AGACACAGGC	AAATTTAGTA	4860
ACTTACCCCG	GGTCACATAG	CTACTGGGTG	GCAAGTTCAG	GGTTAGCTCC	CAGGACAAAT	4920
GCCTCCACAG	CTGGTACTGT	GCTCTGCTTT	ACTGTAGCTA	ATAGTAAAAA	TGGTAGCAAA	4980
AATCAATAGC	AGTAGAACAG	TGCAACAGAT	ATTAAGCGGA	AGAGGAAGAC	TCACAACAAT	5040
GACAACATTT	GTGCTGAAT	TTTTAAGAAC	ACATGGAATT	TCCTTCAGCC	GGGTAGAGAG	5100
AAGATATAGA	AATGTAAACA	CCAAAGATTC	ATAGTTTCTC	TGTATCCCTT	TCAGGGTCAG	5160
TGTGGTTCCCT	GTGGGCTTT	TAGCTCTGTG	GGTGCCCTGG	AGGGCCAAC	CAAGAAGAAA	5220
ACTGGCAAAC	TCTTAAATCT	GAGTCCCCAG	AACCTAGTGG	ATTGTGTGTC	TGAGAATGAT	5280
GGCTGTGGAG	GGGGCTACAT	GACCAATGCC	TTCCAATATG	TGCAGAAGAA	CCGGGGTATT	5340
GACTCTGAAG	ATGCCTACCC	ATATGTGGGA	CAGGTGAGAT	TGCTCCACAC	AATTATACAG	5400
CTCTGTTGGC	TCCTCCTTCC	CCAGCATGAT	GTTTGTACT	GGAAACAATT	CCAGAAATAC	5460
TGTTTCTGT	TATCCTATCC	TGCTTCTTTG	ATGGAATAAT	TTCCACAGA	AGGCCAAGAA	5520
GATTTCACA	ATCTGGGGGA	ATTAGGGAG	CTTAAGCTAC	TATAGCTCCT	ATTGCTATCT	5580
CTGCCATCGA	GAGAAAACAG	AGGCTAGGCT	ACCTACCCCA	TAGACTTCCG	AGCTGGGTTT	5640
TATAACCCCT	TGCTCAATTC	CTCACTCCCA	CAACAAACCC	ACAAACCCAC	CATGCTATTT	5700
TCACAAATG	TGTGGCTTTA	TTTATATGA	TCTCAGTGTG	AGTTTTCAGA	ACATTTTCAGC	5760
AAATATGTA	AGTTTACATG	CTAACATCTA	TAAATGAGA	GAAAAACAA	GTTGCTTCAT	5820

整理番号 156518

ATAAGAGATA AGGGATTAAC TCAGTTCCTC CTGCATGATC CTCTAGTCAT AGGAAGGAAA	5880
TCATATCTGA AAGGGAGGCA ACCTGAGGGG TTTTATATAC ACATAGGGCT GGGTCTGATA	5940
GACAAATATAA TGTAGGGCCT TCACAACAGA AACCTCTGAA ACAGGGACAG CAAGTTTGAG	6000
AATAAAAATG ATGGCTACTG TGTCTAAGC CGTGTCTTA GTGCATTTT TCTTTTCTT	6060
TTTTTCATTT AATCTCATAA CAACTCTGTT AGGTAGACTT ATCTTGAATG TATAGGTGAG	6120
GAAATGACCA CTTAAGGAGA TAAGACAGTA TAATTCATAC CACTAGTATG TAACAATGTA	6180
AGATGTATCT ACCAGGGATG TTTATCTTCT GCAAACATTC CTAGGTATAT CTCCCATGCA	6240
CATGTGCAAG AATTCTTAC TAGGATATAA TGCTTGGA CTGAATGTG TGGGTCTTAG	6300
GGTATGTCTG TCTTCAACTT TACTACACAA TGTCAAATG TTTGCCAAA TATTGGAAA	6360
AATTTATACC TGCAATGTGT AAGAAATCCC CTTCATCAC CTTTTATCA GTATGTTTAT	6420
CTGGCCATTT GCATTTCTTC TTCAGTGAAT TAACTGTTTT TATCTCTTGC TCATTTGTTT	6480
TTCTTTTAT TTTTGTGAAA TAGGGTCTTA CTCTGTGACC CAAGGCTGGA GTGTGGTGAT	6540
ACAGTCATAG CTCACTGCAG CCTCCACTTC CGGGCTCAAG CAATCTCTC GCCTCAGCCT	6600
CCCAAATAGC TAGGATATAG GTGCATGCCA TCATGCCAC CAATTTCAAA AAACCTTTGA	6660
AATTTTTTTT TTGTAAAAGC TAGGCATGGT GGCTCATGCC TGTAAATCCA GCACTTTGGG	6720
AGGCTGAGGT GGGGURCNIN UDAGGATCGC TTGAGCCCAG GAATGGAGG TCGGCCTGAT	6780
ACAACATAGC AAGACCTCAT CTCTACACAA AAAATTTTAA AAGTAGCCA GGTATGATGG	6840
CGTGCATAGT TCTAGCTACT CCGGAAGCTG GTTGGGAGGA CAACTTGAGC CTGGGAGTTC	6900
AAGGCTGCTG TGAACGTGA TCATGTCACT GCTCTCTAGC CTGGGTGACA GAGTGAGACC	6960
CTGTCCCAA AAACAACAAC CGTTTTTTT GGTAGAGACA TTGTCTCGCT ATGTTGCCAA	7020
GGCTAGTCTC AAACCTCTGG GCTCAAGCAA TCCTCCACC TCCCAAAGTG CTGGGATTAT	7080
AGATGTAAGC CACCATGCCT GGCCTACCCT TTTTTTTTT TTTTGAAATG GAAGTTTTC	7140
TTTTGTCAAC TAGGCTTGAG TGCAGTGGC CGATCTTGGC TCACTGCAAC CTCCACCTCC	7200
TGGATTCAG CAATTCCTCT GCCTCAGCCT CCTGAATAGC TGGGATTATA GGCACCCGCA	7260
ACCACGCCCC GCTAGTTTTT GTATTTTTAG TACAGACAG GTTTCACCAT GTTGGCCAGC	7320
TGGTCTTGAA CCCCTGACCT CAGGTGGTCC GCCCGCCTCG GCCTCCCAA ATGCTGGGAT	7380
TAAAGGTGTG AGCCACCATG CCCCACCCCT TACTCATTTT TAATTGGATT GTTTTTCTC	7440
TTTCTTAGCG ATTCTTAAA GTTTAAAGAG AATATTTGGA TACAATACTA TGTATTTAAA	7500
AGTTGAGGTC TGTCTTTCCA TTCTTCTCAC GATGTCTTTC AATCTAGAAA AGTTAATTTT	7560
AATAGGCCTG GCGCGGTGGC TCACGCTTGT AATCCAGCA CTTTGGGAGG CTGAGATGGG	7620
TGGATCACAA GGTCAAGAGA TGAAGACCAT CCTGGCTAAC ATGGGTGAAA CCTGTCTCT	7680
ACTAAAAATA CAAAAAATT AGCTGGGCGT GGTGGCAGGT GCCTGTAGTT CCAGCTACTC	7740
GGGAGGCTGA GGCAGGAGAA TGGCGTGAAC CCGGAGGTG GAGCTTGCAG TGAGCCGAGA	7800
TTGCACCACT GCACTCCAGC CTGGGCAACT GAGCAAGACT GCGTTTCAA AAAAAAAAAA	7860
GTAAATTTTA ATATAGTAAA ATTAGTAAA GGATTAATTT TCCCTTTGCA ATTTTGTAA	7920
TGTGTTTTAT TCGTTTATGA ATGGAGAAAG GTAAGAAAA ATAAATTTA AAAAGAAGA	7980
GATGTGGCCA GGTACGGTGG CTCACACCTA TAATCCCAGT AGTTTGGGAG GCTGAGGCAG	8040
GCAGATCACT TGAGGTCAGG AGTTTGAGAC CAGCTGGGAT AACATGGTGA AACCCCATCT	8100
CTACTAAAAA TACAAAAATT AGCCAGGTGT GATTGCGCAC GCTTGTAATC CCAGCAGGCT	8160
GAGGCAGGAG AATTGCTCGA ACTCAGGAGG CAGAGGTTGC AGTGAGCCAA GATCATGCCA	8220
TTGCACTCCA GCCTGGGTAA CAGAGACTCT GTTCAAAAA AATAAAAAGA TAAAAAGGA	8280
AGAGATCTGA TAGGGCCGCC AGATAACAT TTTAAAGGG ATGGTATTAT AAGTTTGTTC	8340
CCAGCATAAT GCCAGGTAT TCTGACTTTA AAGTATCATC ACATAATATC TTTTGTAGTC	8400
AATTTCCAAG ATATTCTGTT TCACTGTGAA TTCTGTGTA TTTTGGCAC CAGGAGGCAT	8460

整理番号/56518

CAGGGATTG	GAGCACATGG	CAGAAACAAA	GGCATCTTGA	AAAATATCAA	GGCAGTAGAC	8520
CACTGTAATC	TTAAAATGGC	ATATCAAATG	CTGCTATTGC	TGTTAATAT	TAGATAATGT	8580
TAGATAATGT	ATTTTPTTAG	AGGGTATCTC	ACTATCTTGC	ACAGGCTGGA	GTAGAGTGGC	8640
TATTCACAGC	ATGATCACAG	TACACTAAAG	GCTCAAACTC	CTGGGCACAA	ACAATCCTCC	8700
TGCCTCAGCC	TGCTGAGTAG	TAGATAATAA	GTTCTTGTGG	ATGCAACCTT	AGGGTTCTGA	8760
AGGGGTAGTC	TGTAGGAAAA	TGAATTGCTG	AAAAGAATAC	ACCACCTTAA	CATGGGGTAT	8820
TATTCGATTC	CATAATTTGT	GCTTGCCAAT	GAAACATTGC	TAACTACCTG	TAAAATATAG	8880
TGTTGGAAGT	CATAGGCTAA	ATTGCTAAGT	TCTTTAATCT	ATTTTAGTGT	CTTGTTATGT	8940
ACTTTTATAT	TTTGCTTTTG	ATGAGAGCAC	AAGGATCACA	CCAGTTCCCC	TGATATAGGT	9000
GCAGAGGGCC	CAGGTCTTCC	CTCTAGCTAA	GCCTTGGCCT	TGGCCTCCTA	CCCACACAGC	9060
AGCTGGTGCC	TTCTTGCCCC	CTGAGGCTAA	TACATACTAT	GTGGCCAGAA	GATGGTTTAT	9120
GCTTTTAA	AAAATCTTAT	TTCAGAAATC	TTTCCCTACT	GTTTTCTCTC	CACATTTATG	9180
TCTTAAACA	CCTGTAGGGG	ATTTTMTT	TTTTTTTTTT	TTTGAGATGG	AGTCTCGCTC	9240
TGCCCCAGGC	TGGAGTGCAA	TGGCGCGATC	TGGCTCACT	GCAAGGTCTG	CCTCCCAGGT	9300
TCACGCCATT	CTCCTGCCTC	AGCCTCCCCA	GTAGCTGGGA	CTACAGGCGC	CCGCTACCAC	9360
GCCTGGCTAA	TTTTTTTGCA	TTTTTAGTAG	AGACAGGGTT	TCACTGTTAG	CCAGGATGGT	9420
CTCGATCTCC	TGACCTCGTG	ATCCACCCTC	CTCAGCCTCC	AAAGTGCTGG	GATTAACAGG	9480
CATGGAGCCC	CACCGCACTG	GCCTGTATTT	GTGAGGAAGA	ACAGACCCTC	TTTAGAAGCC	9540
CTAGACTGCT	GCCTCTGTTA	GTTCACTGGC	ATCACTCAA	ATATTGGTTG	AGTTTCTTAC	9600
TCACTGAGTT	GGTTTTATG	TGTGGTGGAA	GGGGGAATC	CTCTTTTCAT	ATTGTTCTC	9660
ATTGCCTATT	GCTTTGCTCT	AGTCTATTTA	CAATCTTGTT	TCTTCCAGGA	AGAGAGTTGT	9720
ATGTACAACC	CAACAGGCAA	GGCAGCTAAA	TGCAGAGGGT	ACAGAGAGAT	CCCCGAGGGG	9780
AATGAGAAAG	CCCTGAAGAG	GGCAGTGGCC	CGAGTGGGAC	CTGTCTCTGT	GGCCATTGAT	9840
GCAAGCCTGA	CCTCCTTCCA	GTTTACAGC	AAAGGTAAGA	AGCTGCTGAT	CCTATACAGC	9900
ACTGCTTTTT	ATGATACAAA	CTTGATGGTT	TCTCGAAGGA	CCTTGGGTAT	TTTCAGTACT	9960
TAGTTTTTGT	ATTACATATG	AGGTGGCCAG	AGAGAAATTA	ACAACTGCTG	CAGTATGGAG	10020
CAGCATCTCT	GTGGTAAACC	CTCCTGACAC	GGATGGAATT	CTTCAAACAG	TCTCCTAGAC	10080
TGGGAGATCC	CACAGGGTGA	CCCTTGGATT	GCATAGAGCC	TCACGCTGGT	AGTTTGTATT	10140
CTAGGTGTGT	ATTATGATGA	AAGCTGCAAT	AQCGATAATC	TGAACCATGC	GGTTTGGCA	10200
GTGGGATATG	GAATCCAGAA	GGGAAACAAG	CACTGGATAA	TTAAAAACAG	GTAAATGATGG	10260
GAACACTACT	TTTGTATTTC	AGTCACCCCT	TTAACACTCA	ACCTCACCTC	CAGCTTCCCG	10320
ATATTCTTTT	CTCTGTCCCA	AATCAAGAAA	AAATTATCTC	AGAGTTCTCA	CTTCTATCTT	10380
CTCACTCAGA	GGCTCTTAAT	TCTCACTCTG	ACACTTAATG	GCCAGTGTGT	TAGTCCATTT	10440
TGCATTGCCA	CAAAAGAATA	CCCGAGACTG	GGTAGTTTAT	AAAGAAACGA	GGTTTGTTTG	10500
GCTATACAAA	CGGTGGCACT	AGTATCTGCT	CAGCCTCTGA	TGAGGCCTCA	GAGCTTTTAC	10560
TCATGGCAGA	AGGCAAAAGA	GGGAGCAGGC	ATGTCACATA	GTGAGAGAGG	GAGCAAGAGA	10620
GAGAGGGAGG	TGCCGACTCT	TTAAAGAACC	AGCTCTTGCA	TGAACATAA	GAGTGAGAAC	10680
TCACTCATCA	CCAAGGCGAT	GGCACCAGC	CATTCCATGA	GGAATCCACT	CTCATAACCC	10740
AAACACCTCC	CACATATGCC	CACCTCCCAC	ATTGGGGATC	ACATTTACAG	ATGAGACTGG	10800
GAGGGGACAC	ACATCCAAAC	CATATCCGCC	AGACAATAGT	GCTCAATTAT	GTGCTGGGCA	10860
GATGCTCCCT	GTGTGCAAGG	TGCTTAGTGA	CATACATAAA	CCAACGAGCA	GATGACACCT	10920
TCAGTGAGCT	CAGAGCCCAA	TAAGACAGAC	CTAACTAACC	ATGAGATAAA	GCAGTACAAA	10980
GAACCAGCAG	GAGCTTTGGA	ATTACGTATT	TTTACTTTCT	TTTGTCTCTA	ATGTGATCAG	11040
TTTCTTAGAT	GGTTCCATT	AGCAATCTGT	CTTAAACAGT	AGGGGAGCAG	CGTTAAAGGT	11100

整理番号 156578

TTAATATTCC	TTTTGAACAG	TTTTTTTCCT	TCAAAATACA	CTTAAGATAC	ACGTATATAA	11160
GAACTTGCCA	AAGATTGTGA	AGAGAAACAT	TTTTTAGAAA	TAAGATATAA	ACAAAAAAG	11220
TTAGTGTTAC	TTTCCTATGT	TGGGGAACAA	AGAAACTCC	AGGGTACCTT	GCTTCCCAT	11280
TCTCTTTAGC	ACCTTGTCAC	TTTTGGGGAG	GGGCAGATTG	ATAACAATTA	TAGTTTTCT	11340
TTCTCTGGTG	ATCACCATTA	ACCTGGCAGC	AGCACTGGCT	AAATCTCCTG	TCCTTAGTGC	11400
CCTCCAAGGA	GCAGGAGCCC	TAGACTCTGG	GTCTGCTGACA	GACTCACGCA	GTGGTGTGT	11460
TCAAACCTGA	AGCAACTTTT	TATATCACAG	TTCCAACCTCA	AGGTGAACCT	GAGCATCTTC	11520
CCAAGTCTCC	CACAGCTTCT	GTCTGTGTGT	GTCCCTTCTC	TTGACTCCCA	GGTCCAAGCA	11580
CTTACCCTGT	TCTTTCATGA	TCAGGTACCA	TGTGTGGAGA	TAGCTTCCAA	GAGAGCTGGG	11640
AGGAAGAAAG	GACACACCCG	GGCAGGATCA	GGAACACTGG	GGGCCCCTGG	AGAGGGGGAG	11700
AGTGGGGGAG	GGTACAGGTT	TTAAATAAAA	TGTGTGGTA	ATTAGAGAAT	TGCTGGTTGG	11760
GGAAAGAGGT	CTGAAAACAA	TTCAGGAAGA	TAAACAAGAC	AATCTCTCCT	CTCTCCTCTT	11820
TCTCACGTCG	TCTCTCTTGT	CTTCTAGTCT	CGCTACTCAT	TTCTTAGTA	ATCTCATCCA	11880
CTCTCATAGT	TTTATCCATC	TCTCTATGCG	GGTTTACCCC	CAAATCAAGA	TCACCAGCTT	11940
CAGCTCCTT	CTTATGCTCT	AAACTCACAT	TTTCAAGATT	AATATTCCCC	AAATACAGCT	12000
CTGATCATAT	CACTCTCCCA	CTCAAAATCC	CTCACTGGCT	CCTCACGATG	ATGGGTCACA	12060
GAGTAAAGGT	GAAGCTTTTT	AACCTTGTCAG	TAAAGGTAAT	TCAACCTGAT	CTCAATCTGC	12120
CTTTCAGAC	ATCTCTCCCA	CTACACCTG	TTAGGCACAC	TGCTTTTCAG	CTACATGATC	12180
CTAACAGTGC	CCCACACTTT	CCTGCCCTCTG	TGTTTCATTT	CACACCTTC	CACTGGCATC	12240
CCCTTCCCAC	AGGTCGAAAT	TCTACTTAGC	CTTTTGGCTC	AGCTCAAATG	CCACCTCTTA	12300
CATCAAGCCT	CTAAGATTCT	CTTGATCAGA	AGGAATCTTT	CCCTCTTTTG	ATACCTACAG	12360
TATTATGCTT	TCTCCCTATT	TCTTGACTTT	AAACTCTTTA	AAGTTAAAAA	ACATCATATT	12420
CATTTTTGTG	TACCATCAGT	ACCTCGCACA	ATACTCAGTA	AATATTTTAA	TGAATAAATA	12480
AACTGAGAGT	ACTAAGTATT	TTTCTTGATT	GGTCTTACAG	CTGGGGAGAA	AACTGGGGAA	12540
ACAAAGGATA	TATCTCATG	GCTCGAAATA	AGAACAACGC	CTGTGGCATT	GCCAACTGG	12600
CCAGCTTCCC	CAAGATGTGA	CTCCAGCCAG	CCCAATCCA	TCCTGCTCTT	CCATTTCCTT	12660
CCACGATGGT	GCAGTGTAA	GATGCACTTT	GGAAGGGAGT	TGGTGTGCTA	TTTTTGAAGC	12720
AGATGTGGTG	ATACTGAGAT	TGTCTGTTCA	GTTTCCCAT	TGTTTGTGTC	TTCAAATGAT	12780
CCTTCTTACT	TTGCTTCTCT	CCACCCATGA	CCTTTTTCCT	CTGTGGCCAT	CAGGACTTTC	12840
CCCTGACAGC	TGTGTACTCT	TAGGCTAAGA	GATGTGACTA	CAGCCTGCCC	CTGACTGTGT	12900
TGTCCCAGGG	CTGATGCTGT	ACAGGTACAG	GCTGGAGATT	TTACATAGG	TTAGATTCTC	12960
ATTACGGGA	CTAGTTAGCT	TTAAGCACCC	TAGAGGACTA	GGGTAATCTG	ACTTCTCACT	13020
TCCTAAGTTC	CCTTCTATAT	CCTCAAGGTA	GAAATGTCTA	TGTTTCTTAC	TCCAATTCAT	13080
AAATCTATT	ATAAGTCTTT	GGTACAAGTT	TACATGATAA	AAAGAAATGT	GATTGTCTTT	13140
CCCTTCTTTG	CACCTTTGAA	ATAAAGTATT	TATCTCCTGT	CTACAGTTTA	ATAAATAGCA	13200
TCTAGTACAC	ATTATTTTGT	TGTTGGATAC	TGTGTTAGGT	GCTGGAGGAA	AAAAGATGAA	13260
TAGAACATCT	TCTATGTACT	TGATGCGCTC	ACAGTCTGGT	TGTAGAGACT	GTACATAAAA	13320
CATTTCATCC	CAATTCAATT	ATTGTTTCAT	TCCTTCAGCC	AATATATATT	GAGTTCTTAC	13380
TCTGTGCCAA	GAACTGTACT	ACATTTCTGG	GATTAAAGTG	ATATAAGGAG	ATCTCAGTGT	13440
TTAATCTGCC	TGAGGGGAGA	CTAAATTAAG	TGACATGGAA	ACTTGGGTCT	TGAAAAACAT	13500
TTTAAGGTTA	TTTTTTCTTT	TCTCTCTCTC	TGCTCTCTCT	TTTCTCTCTC	TTTCCTCAGG	13560
GTCTCCCTCT	GTGCCCAGG	CTGGAGTCAG	TGGCACTCAT	AGCTCACTGC	AGCCTTGATC	13620
TCCTGGGCTC	AAGAGTTCTT	CCCACCTCAG	TCTCCTAAGT	AGCTTGGACT	ACGG	13674

整理番号 156578

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1108 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCTTTGGGCTC CCAAAGGCCT GGGATTACAG GCGTGAACCA CTGCGCCTAG CCTGTTAGCA	60
GCTCTTAAAA TCCAGAGGCA TAAGCCTGTA TTTTGGAGGG TTTATGCATG GAATCCAGCT	120
AGAAACTGAG TCTATTACAG ATCCCATTTA TTATCCTTTC TATTCCAAGA AGCCTTTTTT	180
TCTCCTTCCC CACATCTGTT TATGGAAGAA AATGAAGTTT GGGGTGTGGT TTGAGGAATC	240
AGCTAGATTC TTATGATCTG TCACATGCTT GGATGTTGGG GAAGCATTG GAGAAGCTCA	300
TGTGACTTGT CCTAGATTGG GGATTTTAAT TGAGACAGAT GATGTTTATC GGGCATCCCA	360
CCACCTGAGA GTTTTAGCAA CAGAGTCACA TGTGAGTCCA TCAGAACTTA CGGCATTGAT	420
TCAAGTCTG TCATAAATA CCAGGACTGC TGTTTTGGT TACTTTTAAA GACAGTTTCA	480
TCTGGACTTT CTGGGCATAT CCTCCTTCAG CAAAACCACA TTAGGCTGGG AAAACTATTC	540
TGCCTGGAAG TAATGACAAC TTGCAACCAA CAGCTTATA AAAATACAAA GAATTCCTGA	600
GCCTATGGCT TCATTACAT TATTCTTTTA TAGCCTTTA TGTTCAATAC CGCATCCCAG	660
AGGTGAGAGT CAGACACAAA TATGAAAATA GTTTCAATG TTGGAGAGGT AAATCCTAAC	720
AGGAAAGGGG TAGGAAAAGA TATAATCCCC CAATATTAAA ATAAAGATAT TGAAGAAGAA	780
GGATGGGAGA GACTAGGGCT GTGTCCTTCC TTTACTCAC CAAAAGAGAA AGTAAGCTCC	840
TATTTGAGTC AATAGATATT GAGTCTTGT TATTGCCCAC CAAAGACAGT CTTGTGAGAC	900
TAAATAGCTA GTAATTCCCT ACCCTGGCAC ACATGCTGCA TACACACAGA AACACTGCAA	960
ATCCACTGCC TCCTTCCCTC CTCCCTACCC TTCCTTCTCT CAGCATTCT ATCCCCGCT	1020
CCTCCTCTTA CCCAAATTTT CCAGCCGATC ACTGGAGCTG ACTTCCGCAA TCCCGATGGA	1080
ATAAATCTAG CACCCCTGAT GGTGTGCC	1108

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

整理番号 156578

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CACACTTTGC TGCCGAAACG AAGCCAGACA ACAGATTTC ATCAGCAG

48

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1427 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTAACGTTTG CAACTTCCTA GATCTTTTAG CTTTTCATTC CTGTCAATTC TCTGAGTATT	60
AGGGATGTAG TGA CTGAGG ATCACAATAA ACTTTTAGCC TCTGCAGATG AAAACAGAGA	120
TGCACTTCTT AGGTCAATCC CTGGCTAAAT AAAATCTGCC TGGAAATCTG TAGAATTCCT	180
TGTATGATTT ATATATATAC ATACATGATT GTTAGTAAAA GCAAAGTATA TAGGGAATCA	240
TTTCCCCATC CTTCAAGAGT GGCCTTTCTG CAGTGTTTTC TACTTTGGCC AACCAAGGATC	300
AAAACGGTTA ACTCCTTAGT GAGGAGGAGG AGAGTGGTAT GGGGAGGTAG TAGCTCAGTG	360
CTTCCTGTTT ACTGAGACAT CTCAAAGCCC TTAACACTCT AGTTTTTAAA TGTCTACTG	420
GACATTTTGC CAGTTTGCAA AATTACATGT AAATGGACTA TAAGCAATTG TGTAAAGCCAT	480
ATGTCATGCT GCAGGCTGCA AATTGTTCTT AAAATGGAGG ATTTGTAATT AAGAAAGCCA	540
ATGCAAGAAA TGAGTGAAGC TAACTAGAGT AAACCTATGA AAAGCTGTGA ATTTTCATCAT	600
CATAGAACAT TGCTTTTCAG TCTGAACATT CTCTAACAA ACCTTGGATC TGAGGCTTCT	660
TGTCCTTTGC GGCAGCCACA GTGGGTTTTT GTTGTAGGGG GAAAATAAAA AACCTTGCCC	720
GCAGCATCTG GTTAAGATTA GGGCAGTTTC CTGCTTAAGG AGGGAAGGGA GAGAAAAGG	780
AAGAAGAAAT GCATAAGGAG AATGAGGAGA TATACATGT CTCAGAAAC AGGAAACATT	840
GTCTATTTT CCCTTGTCCT CTTCTGACAA GATCTGGGAA AGTACCAGAA TTAGGCACG	900
AAAGAGAAGA ACGCCTCGAA GAAATGATCA GGAAGCAAAA CTTAGACGGA AATCTCTCCT	960
TTGTGTATTC TGAACCCAC TACCACCTG CTATTGTCT GTCTCCAAGC CTGCTAGGGA	1020
CCCTGGAGGA AACGCACTGA GCCCATCTG ATTGTCCAGT TTCTATCCCC CATTTCTGGT	1080

整理番号 156518

```
TGTGTACGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGAGAGAGAG 1140
AGAGACAGAG AGAGAAACAG AGAGAGTGTG TGTGCTTAA ATCTCCCGAG AGAGAGAGAG 1200
AGAGAGAGAG AGAGAGAGAG AGAGAGAAA GAGAGAAATG GCTAAATCCC CCTAGATCAA 1260
ACTCCTTGGA ACCAGATGTA CCAGCATCCT ATCTAAACAC AGGCCCTCC TGAATATCAT 1320
TGTTTATCA CCTTTTTC GTCTACCTT CTCTCCTCA TAAAGCCTAG TTTTCCTCTG 1380
TTTCCCTGCC AATGGAAGA GTTTCCTA ACTACATTCT TCTGCAG 1427
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GATGTGGGG CTCAAGGTTG TGCTGCTACC TGTGGTGAGC TTTGCTCTGT ACCCTGAGGA 60
GATACTGGAC ACCCACTGGG AGCTATGGAA GAAGACCCAC AGGAAGCAAT ATAACAACAA 120
G 121
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 462 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
GTGCCTGGGG TCCTGGAGGG GGCATGGCAG GAAGGCTGAG ACCTGAGCTC TCATCATCTTA 60
GCTTCCAGAC TCCCTTCTTC AATCCAAATG CTTTATTCCA AGCAAATCAG TCCCTCTTCC 120
CTAACTCATG TTAACATACG GTTTTCATTC CTATGCTTCA ATCATCTCT TGTCAAACCT 180
```

整理番号 156518

GTATTCCTTC CCTTTCGTTT TATAAGTGTG TAACATTCCT CTTTGGGAA GAGTCCCAAG	240
ATTAATGCTG TTAATCCATA AGCAATTTT CTGTCTCTCC AGAGCTTGTG TGGTTGTTA	300
CATATTATCT CTCTCTTGC AGGCTCTTAA TTCCATGGTT AGTTCCCCAA CTAAACTGTA	360
AACTTTTATG ATTGTGAGTT TCCTTTATTC TCCTAAAACC CTTACAATA TTACATATGA	420
ACTGTAGACA GTCTATACAA GTACTGACTA TGCTTTGTTT AG	462

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 123 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTGGATGAAA TCTCTCGGCG TTAAATTGG GAAAAAACC TGAAGTATAT TTCCATCCAT	60
AACCTTGAGG CTTCTCTTGG TGTCCATACA TATGAACGG CTATGAACCA CCTGGGGGAC	120
ATG	123

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 85 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCAAGTATAG CTTGAGCTCC TGTCCCACT GCACCATTTG CTTAGTTCC CTGCTGATGC	60
CTGGCCTCTT TCTTCTTTGT CTTAG	85

整理番号 156518

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 156 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACCAGTGAAG AGGTGGTTCA GAAGATGACT GGACTCAAAG TACCCCTGTC TCATTCCCGC	60
AGTAATGACA CCCTTATAT CCCAGAATGG GAAGGTAGAG CCCAGACTC TGTCGACTAT	120
CGAAAGAAAG GATATGTTAC TCCTGTCAA AATCAG	156

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1624 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTACTCTCCT TTCTTCTGGG TGTGCATATG TAATCTGGCA TGACCTTTTC CTTTTTCTGC	60
TGCTTTGTTT TTGAGGTGAA AGGGCACCAG GAAAAGAGGG CAAGGAATTA AGGTACATCT	120
CCCCATTCCC ATTCGTGTTAT TTAACCTCAT TTGTTTCTGT ACATTTCGGT TGTTCCTGGT	180
TTTTCTTTTT CTTTCCCTT TTTTTTTTTT TTTTTTTTTT GAGATAGAGT CTCACTCTGT	240
CGCCCAGGAT GGAGTGCAAT GGTGCAATCT TGGCTCACTG CAACCTACAC CTCCCGGGTT	300
CAAGCGATTC TCCTGCCTCA GCTTCCTGAG TAGCTGAGAT TACAGGCACG CGCCACTACG	360
CTTGGCTAAT TTTTCTATTT TTATAGAGAT GCGTTTTCAC CATGTTGCCC AGGCTGGTCT	420
TGAACGACC TCAGGTGATC CACCTGCCTC AGCCTCCCAA AGTGCTGGGA TTAGAGTCAT	480
GAGCCATCGC GGCCTGGTTT TTCCTTATTA CAAATAGTGT TGCAATAAGC ACCCTTGTGC	540

整理番号 156518

ATATGTTTTT GTGCACATGT ACAAATATTT ATGCAAAATA AGTCCTAAAA TTGGAATTGT	600
TAGGTCACAA ATAATCCTTT CCCCCCCCCC AAATTTTTTT TTTTTTTTTG ACACAGCGTC	660
TCTGTCACCC AGGCTGGAGT CCAGTGGCGC AATCATGGCT CACTGCAGCC TCAACGTCTC	720
AGGCTCAAGT GATTCTCCAA CCTCAGCCTC CCTAGTAGCT GGGAAATTAGA AGCACATGCC	780
ACCACACCCA GCTAATTTTA AAAAATTTTT TGTTAGAGAC AGGGTTTTGC CATGCTACCC	840
AAGCTGGTCT CAAATTCCTG GGCTCAAGCA ATCTGCCCGC TTCGGCCTCC CAAAGTGCTA	900
GGATTACAGA CATGAGCCAC CATGCCCGC CCAAAAAAGT TTTTGCAATC TTACATTCTT	960
ACTAGCATGA GAATGTCAGT TTTTTCACAA CCCAACAAC ACAGGATTGT ATCAGCAAGA	1020
TAAACAATTG ATTAAACGTT CATTTAACAA ACACTTTTTG ACCCCCAGAA CCTACCAGAT	1080
GCAGTGTTAG GCAGCAGAGA CTCAAGATGA CTAAGACACA ACCTGTGTCC TCAGGAAATC	1140
TCAATCTAAA AAAATAGAAC AGGAAGAAGA GAAAAATCTA CAATCTAGCT GCACAAACAA	1200
TAATAGCTAA TACTTTTTGA GATTTTATTG TTTGTCAGGA ACTTCTTAAC TCTTTACATG	1260
AGTTTAAATA TTTAATCCCT TATAACAATA TTTTATGCAT AGAGAACTG AGACACAGGC	1320
AAATTTAGTA ACTTACCCGG GGTACATAG CTACTGGGTG GCAAAGTCAG GGTTAGCTCC	1380
CAGGACAAAT GCCTCCACAG CTGGTACTGT GCTCTGCTTT ACTGTAGCTA ATAGTAAAAA	1440
TGGTAGCAAA AATCAATAGC AGTAGAACAG TGCAACAGAT ATTAAGCGGA AGAGGAAGAC	1500
TCACAACAAAT GACAACATTT GTGCTGAAAT TTTTAAGAAC ACATGGAATT TCCTTCAGCC	1560
GGGTAGAGAG AAGATATAGA AATGTAAACA CCAAAGATTC ATAGTTTCTC TGTATCCCTT	1620
TCAG	1624

(2). INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGTCAGTGTG GTTCCTGTTG GGCTTTTAGC TCTGTGGGTG CCCTGGAGGG CCAACTCAAG	60
AAGAAACTG GCAAACTCTT AAATCTGAGT CCCCAGAACC TAGTGGATTG TGTGTCTGAG	120
AATGATGGCT GTGGAGGGGG CTACATGACC AATGCCTTCC AATATGTGCA GAAGAACCGG	180
GGTATTGACT CTGAAGATGC CTACCCATAT GTGGGACAG	219

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

序列编号 156578

(A) LENGTH: 4326 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i) MOLECULE TYPE: Genomic DNA
(ii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GTGAGATTGC TCCACACAAT TATACAGCTC TGTGGCTCC TCCTTCCCCA GCATGATGTT	60
TTGTACTGGA AACAAATCCA GAAATACTGT TTTCTGTTAT CCTATCCTGC TTTCTTGATG	120
GAATAATTTC CCACAGAAGG CCAAGAAGAT TTCCACAATC TGGGGGAATT TAGGGAGCTT	180
AAGCTACTAT AGCTCCTATT TGCATCTCTG CCATGGAGAG AAAACAGAGG CTAGGGTACC	240
TACCCCATAG ACTTCCGAGC TGGGTCTTAT AACCTCTGTC TCAATTCCTC ACTCCCACAA	300
CAACCCACA AACCCACCAT GCTATTTTCA CAAATTGTGT GGCTTTATTT TATATGATCT	360
CAGTGTGAGT TTTCAGAACA TTTCAGCAAA TTATGTAAGT TTACATGCTA ACATCTATAA	420
AATGAGAGAA AAAACAAGTT GCTTCATATA AGAGATAAGG GATTAAGTCA GTTCCTCCTG	480
CATGATCCTC TAGTCATAGG AAGGAAATCA TATCTGAAAG GGAGGCAACC TGAGGGGTTT	540
TTTATACACA TAGGGCTGGG TCTGATAGAC AATATAATGT AGGGCCTTCA CAACAGAAAC	600
CTCTGAAACA GGGACAGCAA GTTTGAGAAT AAAAATGATG GCTACTGTGT TCTAAGCCGT	660
GTCCTTAGTG CATTTTCTCT TTTCTTTTTT TTCATTTAAT CTCATAACAA CTCTGTTAGG	720
TAGACTTATC TTGAATGTAT AGGTGAGGAA ATGGACACTT AAGGAGATAA GACAGTATAA	780
TTCATACCAC TAGTATGTAA CAATGTAAGA TGTATCTACC AGGGATGTTT ATCTTCTGCA	840
AACATTCCTA GGTATATCTC CCATGCACAT GTGCAAGAAT TTCTTACTAG GATATAATGC	900
CTTGGAACCTG AATTGCTCTG GTCTTAGGGT ATGTCTGTCT TCAACTTTAC TACACAATGT	960
CAAAATGTTT GCCAAAATAT TTGGA AAAAT TTATACCTGC AATGTGTAAG AAATCCCCCTT	1020
CAATCACCTT TTTATCAGTA TGTATCTCTG GCCATTGCA TTTCTTCTTC AGTGAATTAA	1080
CTGTTTTTAT CTCTTGCTCA TTTGTTTTTC TTTTATTTT TTTGAAATAG GGTCTTACTC	1140
TGTTGCCCAA GGCTGGAGTG TGGTGATACA GTCATAGCTC ACTGCAGCCT CCACTTCCGG	1200
GCTCAAGCAA TCCTCTCGCC TCAGCCTCCC AAATAGCTAG GATATAGGTG CATGCCATCA	1260
TGCCCAACAA TTTCAAAAAA CCTTTGAAAT TTTTTTTTTG TAAAAGCTAG GCATGGTGGC	1320
TCATGCCTGT AATCCCAGCA CTTTGGGAGG CTGAGGTGGG AGGATCGCTT GAGCCCAGGA	1380
ATTGGAGGTC GGCCTGATAC AACATAGCAA GACCTCATCT CTACAGAAAA AATTTTTAAA	1440
AGTAGCCAGG TATGATGGCG TGCATAGTTC TAGCTACTCC GGAAGCTGGT TGGGAGGACA	1500
ACTTGAGCCT GGGAGTTCAA GGCTGCTGTG AACTGTGATC ATGTCACTCC TCTCTAGCCT	1560
GGGTGACAGA GTGAGACCTT GTCCCCAAAA ACAACAACCG TTTTTTTTGG TAGAGACATT	1620
GTCTCGCTAT GTTGCCAAGG CTAGTCTCAA ACTCCTGGGC TCAAGCAATC CTCCCACCTC	1680
CCAAAGTGCT GGGATTATAG ATGTAAGCCA CCATGCCTGG CCTACCCCTT TTTTTTTTTT	1740
TTGAAATGGA AGTTTTGCTT TTGTCACCTA GGCTTGAGTG CAGTGGCGCG ATCTTGCTC	1800
ACTGCAACCT CCACCTCCTG GATTCAAGCA ATTCTCCTGC CTCAGCCTCC TGAATAGCTG	1860

整理番号 156518

GGATTATAGG CACCCGCAAC CACGCCCGGC TAGTTTTTGT ATTTTTAGTA CAGACAGGGT	1920
TTCCACCATGT TGGCCAGCTG GTCTTGAACC CCTGACCTCA GGTGGTCCGC CCGCCTCGGC	1980
CTCCCAAAAT GCTGGGATTA AAAGTGTGAG CCACCATGCC CCACCCCTTA CTCATTTTTA	2040
ATTGGATTGT TTTTCTCTT TCTTAGCGAT TCTTAAAAGT TTAAAGAGAA TATTTGGATA	2100
CAACTACTATG TATTTAAAAG TTGAGGTCTG TCTTTCCATT CTTCTCACGA TGTCTTTCAA	2160
TCTAGAAAAG TTAATTTTAA TAGGCCTGGC GCGGTGGCTC ACGCTTGTA TCCCAGCACT	2220
TTGGCAGGCT GAGATGGGTG GATCACAAGG TCAGGAGATG AAGACCATCC TGGCTAACAT	2280
GGGTGAAACC CTGTTTCTAC TAAAAATACA AAAAAATTAG CTGGGCGTGG TGGCAGGTGC	2340
CTGTAGTTCC AGCTACTCGG GAGGCTGAGG CAGGAGAATG GCGTGAACCC GGGAGGTGGA	2400
GCITGCAGTG AGCCGAGATT GCACCACTGC ACTCCAGCCT GGGCAACTGA GCAAGACTGC	2460
GTTTCAAAA AAAAAAAGT TAATTTAAT ATAGTAAAT TAGTAAAGG ATTAATTTTC	2520
CCTTTGCAAT TTTTGTAAATG TGTTTTATTC GTTTATGAAT GGAGAAAGGT AAGAAAAAT	2580
AAAATTTAA AAAGAAGAGA TGTGGCCAGG TACGGTGGCT CACACCTAT ATCCCAGTAG	2640
TTTGGGAGGC TGAGGCAGGC AGATCACTTG AGGTGAGGAG TTTGAGACCA GCTGGGATAA	2700
CATGGTGAAA CCCCATCTCT ACTAAAAATA CAAAAATTAG CCAGGTGTGA TTGCGCACGC	2760
TTGTAATCCC AGCAGGCTGA GGCAGGAGAA TTGCTCGAAC TCAGGAGGCA GAGGTTGCAG	2820
TGAGCCAAGA TCATGCCATT GCACTCCAGC CTGGGTAAAC GAGACTCTGT TTCAAAAAA	2880
TAAAAAGATA AAAAAGGAAG AGATCTGATA GGGCGGCCAG ATAAACATTT TAAAGGGGAT	2940
GCTATTATAA GTTTGTTCCTC AGCATAATGC CAGGTTATTC TGACTTTAAA GTATCATCAC	3000
ATAATATCTT TTTGAGTCAA TTTCCAAGAT ATTCTGTTTC ACTTGTAATT CTGTGTAATT	3060
TTTGGCACCA GGAGGCATCA GGGATTGGA GCACATGGCA GAAACAAAG CATCTTGAAA	3120
AATATCAAGG CAGTAGACCA CTGTAATCTT AAAATGGCAT ATCAAATGCT GCTATTGCTG	3180
TTAATATTTA GATAATGTTA GATAATGTAT TTTTTTAGAG GGTATCTCAC TATCTTGCAC	3240
AGGCTGGAGT AGAGTGGCTA TTCACAGCAT GATCAGTA CACTAAAGGC TCAAACCTCT	3300
GGGCACAAAC AATCCTCCTG CCTCAGCCTG CTGAGTAGTA GATAATAAGT TCTTGTGGAT	3360
GCAACCTTAG GGTCTGAAG GGGTAGTCTG TAGGAAAATG AATTGCTGAA AAGAATACAC	3420
CACCTTAACA TGGCCTATTA TTCGATTCCA TAATTGTGGC TTGCCAATGA AACATTGCTA	3480
ACTACCTGTA AAATATAGTG TTGGAAGTCA TAGGCTAAAT TGCTAAGTTC TTTARTCTAT	3540
TTTAGTGTCT TGTTATGTAC TTTTATATTT TGTCTTTGAT GAGAGCACAA GGATCACACC	3600
AGTTCCCTCG ATATAGGTGC AGAGGGCCCA GGTCTTCCCT CTAGCTAAGC CTTGGCCTTG	3660
GCCTCCTACC CACACAGCAG CTGGTGCCCT CCTGCCCCCT GAGGCTAATA CATACTATGT	3720
GGCCAGAAGA TGGTTTATGC TTTTAAAAA AATCTTATTT CAGAAATCTT TCCCTACTGT	3780
TTTCTCCCA CATTTATGTC TTAACACACC TGTAGGGGAT TTTTMTTTT TTTTMTTTT	3840
TGAGATGGAG TCTCGCTCTC GCCCAGGCTG GAGTGCAATG GCGCGATCTT GGCTCACTGC	3900
AAGGTCTGCC TCCCAGGTTT ACGCCATTCT CCTGCCTCAG CCTCCCCAGT AGCTGGGACT	3960
ACAGGCGCCC GCTACCACGC CTGGCTAATT TTTTGCATT TTTAGTAGAG ACAGGTTTC	4020
ACTGTTAGCC AGGATGGTCT CGATCTCCTG ACCTCGTGAT CCACCTCCT CAGCCTCCAA	4080
AGTGCTGGGA TTAACAGGCA TGGAGCCCCA CCGCACTGGC CTGTATTTGT GAGGAAGAAC	4140
AGACCTCTT TAGAAGCCCT AGACTGCTGC CTCTGTTAGT TCACTGGCAT CACTCAAAAT	4200
ATTGGTTGAG TTTCTTACTC ACTGAGTTGG TTTTATGTG TGGTGAAGG CGGGAATCCT	4260
CTTTTCATAT TCGTCTCAT TGCTATTCG TTTGTCCTAG TCCTATTACA ATCTTGTTTC	4320
TTCCAG	4326

整理番号 15651D

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 166 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
GAAGAGAATT GTATGTACAA CCCAACAGGC AAGGCAGCTA AATGCAGAGG GTACAGAGAG      60
ATCCCCGAGG GGAATGAGAA AGCCCTGAAG AGGGCAGTGG CCCGAGTGGG ACCTGTCTCT      120
GTGGCCATTG ATGCAAGCCT GACCTCCTTC CAGTTTACA GCAAAG                      166
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 270 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
GTAAGAAGCT GCTGATCCTA TACAGCACTG TCTTTATGA TACAAACTTG ATGGTTTCTC      60
GAAGGACCTT GGGTATTTTC AGTACTTAGT TTTTGTATTC ACATGGAGGT GGGCAGAGAG      120
AAATTAACAA CTGCTGCAGT ATGGAGCAGC ATCTCTGTGG TAAACCCCTCC TGACACGGAT      180
GGAATTCTTC AAACAGTCTC CTAGACTGGG AGATCCCAACA GGGTGACCCT TGGATTGCAT      240
AGAGCCTCAC GCTGGTAGTT TGTATTCTAG                      270
```

(2) INFORMATION FOR SEQ ID NO:15:

整理番号 156518

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 106 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
GTGTGTATTA TGATGAAAGC TGCAATAGCG ATAATCTGAA CCATGCGGTT TTGGCAGTGG    60
GATATGGAAT CCAGAAGGGA AACAAAGCACT GGATAATTAA AAACAG                      106
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2270 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
GTAATGATGG GAACACTACT TTGTTATTTC AGTCACCCCTT TTAACACTCA ACCTCACCTC    60
CAGCTTCCCG ATATTCCTTT CTCTGTCCCA AATCAAGAAA AAATTATCTC AGAGTTCTCA    120
CTTCTATCTT CTCAGTCAGA GGCTCTTAAT TCTCAGTCTG ACACTTAATG GCCAGTGTOT    180
TAGTCCATTT TGCAATTGCCA CAAAAGAATA CCCGAGACTG GGTAGTTTAT AAAGAAACGA    240
GGTTTGTTTG GCTATACAAA GCGTGGCACT AGTATCTGCT CAGCCTCTGA TGAGGCCTCA    300
GAGCTTTTAC TCATGGCAGA AGGCAAAAGA GGGAGCAGGC ATGTACATA GTGAGAGAGG    360
GAGCAAGAGA GAGAGGGAGG TGCGACTCT TTAAGAACC AGCTCTTGCA TGAATAATA    420
GAGTGAGAAC TCACTCATCA CCAAGGCGAT GGCACCAAGC CATTCCATGA GGAATCCACT    480
CTCATAACCC AACACCTCC CACTATGCCC CACCTCCAC ATTGGGGATC ACATTTGAGC    540
ATGAGACTGG GAGGGGACAC ACATCCAAAC CATATCCGCC AGACAATAGT GCTCAATTAT    600
GTGCTGGGCA GATGCTCCCT GTGTGCAAGG TGCTTAGTGA CATACATAAA CCAACGAGCA    660
```

整理番号 156518

GATGACACCT TCAGTGAGCT CAGAGCCCAA TAAGACAGAC CTAACCTAACC ATGAGATAAA	720
GCAGTACAAA GAACCAGCAG GAGCTTTGGA ATTACGTATT TTTACTTTCT TTTGTCCTTA	780
ATGTGATCAG TTTCTTAGAT GGTTCCTATT AGCAATCTGT CTTTAACAGT AGGGGAGCAG	840
CGTTAAAGGT TTAATATTCC TTTTGAACAG TTTTTCCT TCAAAATACA CTTAAGATAC	900
ACGTATATAA GAACCTGCCA AAGATTGTGA AGAGAAACAT TTTTGAATAA TAAGATATAA	960
ACAAAAAAG TTAGTGTTAC TTTCTATGT TGGGAACAA AGAAACTCC AGGGTACCTT	1020
GCTTCCCAT TCTCTTAGC ACCTTGTGAC TTTTGGGAG GGGCAGATTG ATAACAATTA	1080
TAGTTTTCCT TTCTGGCTG ATCACCATTA ACCTGGCAGC AGCACTGGCT AAATCTCCTG	1140
TCCTTAGTGC CCTCCAAGGA GCAGGAGCCC TAGACTCTGG GTCGCTGACA GACTCACGCA	1200
GTGGTGTGT TCAAACCTGA AGCAACTTTT TATATCACAG TTCCAACCTA AGGTGAACCT	1260
GAGCATCTTC CCAAGTCTCC CACAGCTTCT GTCTGTGTT GTCCCTTCTC TTGACTCCCA	1320
GGTCCAAGCA CTTACCCTGT TCTTTCATGA TCAGGTACCA TGTGTGGAGA TAGCTTCCAA	1380
GAGAGCTGGG AGGAAGAAAG GACACACCCG GGCAGGATCA GGAACACTGG GGGCCCCCTG	1440
AGAAGGGGAG AGTGGGGGAG GGTACAGGTT TTAATAAAA TGTGTGGTA ATTAGAGAAT	1500
TGCTGGTGG GGAAGAGGT CTGAAAACA TTCAGGAAGA TAAACAAGAC AATCTCTCCT	1560
CTCTCCTCT TCTCACGTCG TCTCTCTGT CTCTAGTCT CGCTACTCAT TTCCTTAGTA	1620
ATCTCATCCA CTCTCATAGT TTCATCCATC TCTCTATGG GGTTTACCCC CAAATCAAGA	1680
TCACCAGCTT CAGCCTCCTT CTTATGCTCT AAACCTACAT TTTCAAGATT AATATTCCTC	1740
AAATACAGCT CTGATCATAT CACTCTCCA CTCAAAATCC CTCCTGGCT CCTCACGATG	1800
ATGGGTGACA GAGTAAAGGT GAAGCTTTT AACCTTGCAG TAAAGGTAAT TCAACCTGAT	1860
CTCAATCTGC CTTTCAGAC ATCTCTCCA CTACACCTG TTAGGCACAC TGCTTTTCAG	1920
CTACATGATC CTAACAGTGC CCCACACTTT CCTGCCTCTG TTCTTCATTT CACACCCTTC	1980
CACTGGCATC CCCTTCCCAC AGGTCGAAAT TCTACTTAGC CTTTGGCTC AGCTCAAATG	2040
CCACCTCTTA CATCAAGCCT CTAAGATTCT CTTGATCAGA AGGAATCTTT CCTCCTTTG	2100
ATACCTACAG TATTATGCCT TCTCCCTATT TCTTGACTTT AAACCTTTTA AAGTTAAAAA	2160
ACATCATATT CATTTTGTG TACCATCAGT ACCTCGACA ATACTCAGTA AATATTTTAA	2220
TGAATAAATA AACTGAGAGT ACTAAGTATT TTTCTTGATT GGTCTTACAG	2270

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 97 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

处理序号 / 56518

CTGGGGAGAA AACTGGGGAA ACAAGGATA TATCCTCATG GCTCGAAATA AGAACAACGC 60
CTGTGGCATT GCCAACCTGG CCAGCTTCCC CAAGATG 97

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 598 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TOACTCCAGC CAGCCCAAAT CCATCCTGCT CTTCCATTTC CTTCCACGAT GGTGCAGTGT	60
AACGATGCAC TTTGGAAGGG AGTTGGTGTG CTATTTTGA AGCAGATGTG GTGATACTGA	120
GATTGTCTGT TCAGTTTCCC CATTGTGTTG TGCTTCAAAT GATCCTTCCT ACTTTGCTTC	180
TCTCCACCCA TGACCTTTT CCACGTGTGGC CATCAGGACT TTCCCCTGAC AGCTGTGTAC	240
TCTTAGGCTA AGAGATGTGA CTACAGCCTG CCCCCTGACTG TGTGTCCCA GGGCTGATGC	300
TGTACAGGTA CAGGCTGGAG ATTTTCACAT AGGTTAGATT CTCATTCACG GGACTAGTTA	360
GCTTTAAGCA CCCTAGAGGA CTAGGGTAAT CTGACTTCTC ACTTCCTAAG TTCCCTTCTA	420
TATCCTCAAG GTAGAAATGT CTATGTTTTT TACTCCAATT CATAAATCTA TTCATAAGTC	480
TTTGGTACAA GTTTACATGA TAAAAAGAAA TGTGATTGTG CTTCCCTTCT TTGCACTTTT	540
GAAATAAAGT ATTTATCTCC TGTCTACAGT TTAATAAATA GCATCTAGTA CACATTCA	598

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 459 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

处理番号 156578

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTTTGTGTTG GATACTGTGT TAGGTGCTGG AGGAAAAAAG ATGAATAGAA CATCTTCTAT	60
GTACTTCATG CGCTCACAGT CTGGTTGTAG AGACTGTCAC ATAAACATTT CATCCCAATT	120
CATTTATPTG TTCATTCCTT CAGCCAATAT ATATTGAAGT CTTACTCTGT GCCAAGAACT	180
GTACTACATT TCTGGGATTA AGTGGATATA AGGAGATCTC AGTGTTTAAT CTGCCTGAGG	240
GGAGACTAAA TTAAGTGACA TGGAAACTTG GGTCTTGAAA AACATTTTAA GGTTATTTT	300
TCCTTTCTCT CTCTCTCGCT CTGCTTTTCT CTCTCTTTCTG TCAGGGTCTC CCTCTGTTGC	360
CCAGGCTGGA GTCAGTGGCA CTCATAGCTC ACTGCAGCCT TGATCTCCTG GGCTCAAGAG	420
TTCTTCCAC CTCAGTCTCC TAAGTAGCTT GGACTACGG	459

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 329 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Trp Gly Leu Lys Val Leu Leu Leu Pro Val Val Ser Phe Ala Leu	
1 5 10 15	
Tyr Pro Glu Glu Ile Leu Asp Thr His Trp Glu Leu Trp Lys Lys Thr	
20 25 30	
His Arg Lys Gln Tyr Asn Asn Lys Val Asp Glu Ile Ser Arg Arg Leu	
35 40 45	
Ile Trp Glu Lys Asn Leu Lys Tyr Ile Ser Ile His Asn Leu Glu Ala	
50 55 60	
Ser Leu Gly Val His Thr Tyr Glu Leu Ala Met Asn His Leu Gly Asp	
65 70 75 80	
Met Thr Ser Glu Glu Val Val Gln Lys Met Thr Gly Leu Lys Val Pro	
85 90 95	
Leu Ser His Ser Arg Ser Asn Asp Thr Leu Tyr Ile Pro Glu Trp Glu	
100 105 110	
Gly Arg Ala Pro Asp Ser Val Asp Tyr Arg Lys Lys Gly Tyr Val Thr	
115 120 125	

75

整理番号 156518

```

Pro Val Lys Asn Gln Gly Gln Cys Gly Ser Cys Trp Ala Phe Ser Ser
130                      135                      140
Val Gly Ala Leu Glu Gly Gln Leu Lys Lys Lys Thr Gly Lys Leu Leu
145                      150                      155                      160
Asn Leu Ser Pro Gln Asn Leu Val Asp Cys Val Ser Glu Asn Asp Gly
165                      170                      175
Cys Gly Gly Gly Tyr Met Thr Asn Ala Phe Gln Tyr Val Gln Lys Asn
180                      185                      190
Arg Gly Ile Asp Ser Glu Asp Ala Tyr Pro Tyr Val Gly Gln Glu Glu
195                      200                      205
Ser Cys Met Tyr Asn Pro Thr Gly Lys Ala Ala Lys Cys Arg Gly Tyr
210                      215                      220
Arg Glu Ile Pro Glu Gly Asn Glu Lys Ala Leu Lys Arg Ala Val Ala
225                      230                      235                      240
Arg Val Gly Pro Val Ser Val Ala Ile Asp Ala Ser Leu Thr Ser Phe
245                      250                      255
Gln Phe Tyr Ser Lys Gly Val Tyr Tyr Asp Glu Ser Cys Asn Ser Asp
260                      265                      270
Asn Leu Asn His Ala Val Leu Ala Val Gly Tyr Gly Ile Gln Lys Gly
275                      280                      285
Asn Lys His Trp Ile Ile Lys Asn Ser Trp Gly Glu Asn Trp Gly Asn
290                      295                      300
Lys Gly Tyr Ile Leu Met Ala Arg Asn Lys Asn Asn Ala Cys Gly Ile
305                      310                      315                      320
Ala Asn Leu Ala Ser Phe Pro Lys Met
325

```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCGAAACGAA GCCAGACAAC

20

管理番号 156518

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTCACCACA GGTAGCAGCA C

21

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CTGCTGCTAC CTGTGGTGAG C

21

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

77

整理 号 156518

- (ii) MOLECULE TYPE: Genomic DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCCAAATTAA ACGCCGAGAG

20

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CTCTCGGCGT TTAATTGGG

20

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

78

整理番号 156518

GGTACTTTGA GTCCAGTCAT C

21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CCAGACTCTG TCGACTATCG

20

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CACATATGGG TAGGCATCTT C

21

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

79

整理番号 156578

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GAAGATGCCT ACCCATATGT G

21

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTTACATTAT CGCTATTGCA C

21

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

80

整理番号 156518

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAAAGGTGT GTATTATGAT G

21

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCCGTGTGTC TTATTTCGAG C

21

What is claimed is:

1. An isolated polynucleotide comprising a region selected from the group consisting of:

- 5 a sequence at least 80% identical to the sequence in SEQ ID NO: 1,
 a sequence at least 85% identical to the sequence in SEQ ID NO: 1,
 a sequence at least 90% identical to the sequence in SEQ ID NO: 1,
 a sequence at least 95% identical to the sequence in SEQ ID NO: 1, and
10 a sequence at least 97% identical to the sequence in SEQ ID NO: 1.

10 2. An isolated polynucleotide according to claim 1, wherein said region is a genomic DNA or a cDNA.

15 3. An isolated polynucleotide comprising cathepsin K enhancer or promoter.

 4. The isolated polynucleotide of claim 3 having the sequence in SEQ ID NO: 1.

 5. An isolated polynucleotide comprising cathepsin K polyadenylation region.

20 6. The isolated polynucleotide of claim 5 having the sequence in SEQ ID NO: 1.

 7. An isolated polynucleotide comprising a cathepsin K intron.

25 8. The isolated polynucleotide of claim 7 having the sequence in SEQ ID NO: 1.

 9. An isolated polynucleotide comprising a sequence selected from the group consisting of:

 intron 1, 2, 3, 4, 5, 6 and 7.

30 10. An isolated polypeptide encoded by a polynucleotide comprising a sequence selected from the group consisting of:

 intron 1, 2, 3, 4, 5, 6 and 7.

11. An isolated polynucleotide comprising a cathepsin K exon.
12. The isolated polynucleotide of claim 11 having the sequence in SEQ ID NO: 1.
- 5 13. An isolated polynucleotide comprising a sequence selected from the group consisting of:
- exon 1, 2, 3, 4, 5, 6, 7 and 8.
- 10 14. An isolated polypeptide encoded by a polynucleotide comprising a sequence selected from the group consisting of:
- exon 1, 2, 3, 4, 5, 6, 7, and 8.
- 15 15. An isolated polynucleotide comprising an exon-exon pairs selected from the group consisting of:
- 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 2-4, 2-5, 2-6, 2-7, 2-8, 3-4, 3-5, 3-6, 3-7, 3-8, 4-5, 4-6, 4-7, 4-8, 5-7, 5-8 and 6-8.
- 20 16. An isolated polypeptide encoded by a polynucleotide comprising an exon-exon pairs selected from the group consisting of:
- 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 2-4, 2-5, 2-6, 2-7, 2-8, 3-4, 3-5, 3-6, 3-7, 3-8, 4-5, 4-6, 4-7, 4-8, 5-7, 5-8 and 6-8.
- 25 17. An isolated polynucleotide comprising a region selected from the group consisting of:
- a sequence at least 80% identical to the human cDNA in ATCC Deposit No.: 98035,
- 30 98035,
- a sequence at least 85% identical to the human cDNA in ATCC Deposit No.: 98035,
- 98035,
- a sequence at least 90% identical to the human cDNA in ATCC Deposit No.: 98035,
- a sequence at least 95% identical to the human cDNA in ATCC Deposit No.: 98035, and

a sequence at least 97% identical to the human cDNA in ATCC Deposit No.: 98035.

18. An isolated polynucleotide comprising a member selected from the group
5 consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20;

(b) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising amino acid 1 to amino acid 329 forth in SEQ ID NO: 20;

10 (c) a polynucleotide which hybridizes to and is at least 70% complementary to the polynucleotide of (a) or (b); and

(d) a polynucleotide fragment of the polynucleotide of (a), (b) or (c) wherein said fragment comprises at least 30 consecutive bases of SEQ ID NO: 1.

15 19. An isolated polynucleotide according to claim 1, wherein said region is the region encoding cathepsin K in the human cDNA insert in ATCC Deposit No.: 98035.

20 20. An expression vector comprising cis-acting control elements effective for expression in a host cell of an operatively linked polynucleotide according to claim 1.

21. An expression vector according to claim 20, wherein said control elements are effective for inducible expression of said polynucleotide in said host cell.

25 22. An expression vector comprising cis-acting control elements effective for expression in a host cell of an operatively linked polynucleotide according to claim 18.

23. A host cell having expressibly incorporated therein an expression vector according to claim 22.

30 24. A host cell having expressibly incorporated therein an expression vector according to claim 18.

25. A process for making a polypeptide, comprising the step of expressing in a host cell a polynucleotide according to claim 1.

26. A process for making a polypeptide, comprising the step of expressing in a host cell a polynucleotide according to claim 18.

5 27. A polypeptide of 15 or more amino acids identical in sequence to a continuous region of the amino acid sequence of SEQ ID NO: 20 .

28. A polypeptide of 15 or more amino acids identical in sequence to a continuous region of the amino acid sequence of the polypeptide encoded by the human cDNA in
10 ATCC Deposit No. 98035.

29. A polypeptide of 50 or more amino acids identical in sequence to a continuous region of the amino acid sequence of SEQ ID NO: 20.

15 30. A polynucleotide of 25 or more nucleotides identical in sequence to a continuous region of a polynucleotide at least 90% identical in sequence to the polynucleotide of SEQ ID NO: 1.

31. A polynucleotide according to claim 23 of 50 or more nucleotides.
20

32. A polynucleotide according to claim 31 of 75 or more nucleotides.

33. A host cell genetically engineered with the vector of Claim 20.

25 34. A process for producing a polypeptide comprising: expressing from the host cell of Claim 33 the polypeptide encoded by said DNA.

35. A process for producing cells capable of expressing a polypeptide comprising genetically engineering cells with the vector of Claim 20.

30 36. A method for determining a cathepsin K-encoding polynucleotide in a sample, comprising the steps of:
 hybridizing to a sample a probe specific for said polynucleotide under conditions effective for said probe to hybridize specifically to said polynucleotide and

determining the hybridization of said probe to polynucleotides in said sample,
wherein said probe comprises its sequence a region of 20 or more base pairs at least 90%
identical to the polynucleotide sequence of SEQ ID NO: 1.

37. A method for determining a cathepsin K-encoding polynucleotide in a sample,
5 comprising the steps of:

hybridizing to a sample a probe specific for said polynucleotide under conditions
effective for said probe to hybridize specifically to said polynucleotide and

determining the hybridization of said probe to polynucleotides in said sample,
wherein said probe comprises its sequence a region of 20 or more base pairs at least 90%
10 identical to the polynucleotide sequence of the human cDNA insert of ATCC deposit No.
98035.

38. A method for detecting in a sample a polypeptide comprising a region at least
90% identical in sequence to the amino acid sequence of residues 1 through 135 of SEQ ID
15 NO: 1, said method comprising:

incubating with a sample a reagent that binds specifically to said polypeptide under
conditions effective for specific binding and

determining the binding of said reagent to said polypeptide in said sample.

39. A method for diagnosing a disease characterized by aberrant expression of a
cathepsin K polynucleotide, comprising:

hybridizing a probe specific for a polynucleotide comprising a region at least 90%
identical in sequence to an RNA or DNA that encodes amino acids 1 - 329 in SEQ ID NO:
1 under condition effective for specific hybridization and

25 determining hybridization of said probe to said polynucleotide in said sample.

40. A method for diagnosing a disease characterized by aberrant expression of a
cathepsin K polypeptide, comprising:

incubating with a sample a reagent that binds specifically to a polypeptide
30 comprising a region at least 90% identical in sequence to the amino acid sequence of
residues 1 through 329 of SEQ ID NO: 1 under conditions effective for specific binding,
and

determining the binding of said reagent to said polypeptide in said sample.

41. A compound which inhibits activation of the polypeptide of claim 13.
42. A method for the treatment of a patient having need of cathepsin K comprising:
administering to the patient a therapeutically effective amount of the polypeptide of claim
5 13.
43. The method of Claim 42 wherein said therapeutically effective amount of the
polypeptide is administered by providing to the patient DNA encoding said polypeptide and
expressing said polypeptide *in vivo*.
10
44. A method for the treatment of a patient having need to inhibit a cathepsin K
polypeptide comprising: administering to the patient a therapeutically effective amount of
the compound of Claim 41.
45. A process for diagnosing a disease or a susceptibility to a disease related to an
15 under-expression of the polypeptide of claim 13 comprising:
determining a mutation in a nucleic acid sequence encoding said polypeptide.
46. A diagnostic process comprising:
20 analyzing for the presence of the polypeptide of claim 13 in a sample derived from
a host.
47. A method for identifying compounds which bind to and inhibit activation of
the polypeptide of claim 13 comprising:
25 contacting a cell expressing on the surface thereof a receptor for the polypeptide,
said receptor being associated with a second component capable of providing a detectable
signal in response to the binding of a compound to said receptor, with an analytically
detectable cathepsin K polypeptide and a compound under conditions to permit binding to
the receptor; and
30 determining whether the compound binds to and inhibits the receptor by detecting
the absence of a signal generated from the interaction of the cathepsin K with the receptor.

FIGURE 1

[SEQ ID NO. 1]

GCTTTGGCTCCCAAAGGCCTGGGATTACAGGCGTGAACCACTGCGCCTAG
CCTGTTAGCAGCTCTTAAATCCAGAGGCATAAGCCTGTATTTTGGAGG
TTTATGCATGGAATCCAGCTAGAACTGAGTCTATTACAGATCCCATTTA
TTATCCTTTCTATTCCAAGAAGCCTTTTTTCTCCTTCCCCACATCTGTT
TATGGAAGAAAATGAAGTTTGGGGTGTGGTTTGAGGAATCAGCTAGATT
TTATGATCTGTACATGCTTGGATGTTGGGGAAGCATTGGAGAAGCTCA
TGTGACTTGTCTAGATTGGGGATTTTAAATGAGACAGATGATGTTTATC
GGGCATCCCACCACCTGAGAGTTTACCAACAGAGTCACATGTGACTCCA
TCAGAACTTACGGCATTGATTCAAGTGTCTCATATAAATAACAGGACTGC
TGTTTTTGGTTACTTTTAAAGACAGTTTCATCTGGACTTTCTGGGCATAT
CCTCCTTCAGCAAAACCACATTAGGCTGGGAAAATATTCTGCCTGGAAG
TAATGACAACTTGCAACCAACAAGCTTATAAAAAATCAAAAGAAATCTGGA
GCCTATGGCTTCCATTACATTATTCTTTTATAGCCTTTTATGTTTCATTAC
CGCATCCCAGAGGTGAGAGTCAGACACAAATATGAAAATAGGTTTCAATG
TTGGAGAGGTAAATCCTAACAGGAAAGGGTAGGAAAAGATATAATCCCC
CAATATTAAAAATAAGATATTGAAGAAGAAGGATGGGAGAGACTAGGGCT
GTGTCTTCTCTTTTACTCACCAAAAGAGAAAGTAAGCTCTATTGAGTC
AATAGATATTGAGGTCTTGTATTGTCACCAAGACAGTCTTGTGAGAC
TAAATAGCTAGTAATTCCTTACCCTGGCACACATGCTGCATACACACAGA
AACACTGCAATCCACTGCCCTCCTTCCCTCCTCCTTACCCTTCTTCTCT
CAGCATTTCTATCCCCGCCCTCCTCCTTACCCTTCTTCCAGCCGATC
ACTGGAGCTGACTTCCGCAATCCCGATGGAATAAATCTAGCACCCCTGAT
GGTGTGCCACACTTTGCTGCCGAAACGAAGCCAGACAACAGATTTCAT
CAGCAGgtaacgtttg caacttccta gatcttttag cttttcattcc
tgtcaattctctgagtattagggatgtagtgcacttgaggatcacataaa
cttttagcctctgcagatgaaaacagagatgcacttcttaggtcattccc
tggctaaataaaatctgcctggaaaatctgtagaattccttgtatgattta
tatataacatcatgatgttagtaaaagcaaagtatatagggatcat
tcccccatccttcaagagtggccttctgcagtgttttctactttggcca
acaaggatcaaaacggttaactccttagtgaggaggaggagagtggatg
gggaggtagtagctcagtgcttccctgttccactgagacatctcaaagccct
taacactctagtttttaaaatgtcctactggacattttgccagtttgcaa
attacatgtaaatggactataagcaattgtgtgaagccatgtcatgtctg
caggctgcaaaatgttctttaaattggaggatttgaattaaagaaagccaa
tgcaagaaatgagtgaaagctaactagagtaaaacttatgaaaagctgtgaa
tttcatcatcatagaaacattgtctttcagtcctgaacattcttctaacaaa
ccttggatctgaggcttcttgtccttttgcggcagccacagtggtttttg
ttgttaggggaaaaataaaaaaccttgcggcagcatctggttaagattag
ggcagtttctctgcctaaggagggagggagagaaaaaggaagaagaatg
cataaggagaatgaggagatatcaaatgtctcagaaaaacaggaaacattg
tcctattttcccttgtccttctctgacaagatctgggaaagtaccagaat
ttaggcacgaaagagaagaacgcctcgaagaaatgatcaggaagcaaaac
tagacggaaatctctccttctgtgtattctgaacccactaccaccttgc
tatttgtctg tctccaagcc tgcaggagac cctggaggaa acgactgag
cccatctga ttgtccagtt tctatcccc atttctgggt gtgtacgtgt
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gagagagaga

FIGURE 1A

gagacagaga gagaaacaga gagagtgtgt gttgcctaaa tctcccgaga
 gagagagaga gagagagaga gagagagaga gagagaaaag agagaaatgg
 ctaaatcccc ctagatcaaa gtccttggaa ccagatgtac cagcatccta
 tctaaacaca ggccctcct gactatcatt gttttatcac cctttttccg
 cctacccttc ccttccctcat aaagcctagt tttcctctgt ttccctgcca
 aatggaagag ttttccctaa ctacattctt ctgcagGATG TGGGGGCTCA
AGGTTCTGCT GCTACCTGTG GTGAGCTTTG CTCTGTACCC TGAGGAGATA
CTGGACACCC ACTGGGAGCT ATGGAAGAAG ACCCACAGGA AGCAATATAA
 CAACAAGgtg cctgggggtcc tggagggggc atggcaggaa ggcctgag
 acctgagctc tctcatctta gcttcagac tcccttcttc aatccaaatg
 ctttattcca agcaaatcag tccctcttcc ctaactcatg ttaacatagc
 gttttcattc ctatgcttca atcatcctct tgtcaaactt gtattccttc
 cctttgggtt tataagtgtg taacattcct cttttgggaa gagtcccaag
 attaatgctg ttaatccata agcaattttt ctgtctctcc agagcttgtg
 tgggtgttta catattatct ctcttcttgc aggtctctaa ttccatgggt
 agttccccaa ctaaactgta aacttttatg attgtgagtt tcccttattc
 tccataaacc cttcacaata ttacatatga actgtagaca gtctatacaa
 gtactgactatgctttgttttagGTGGATGAAATCTCTCGGCGTTTAAATTGGGA
AAAAAACCTGAAGTATATTTCCATCCATAACCTTGAGGCTTCTCTTGGTGTC
TACATATGAAGTGGCTATGAACACCTGGGGGACATGgcaagtatagcttcagc
 tccgtgtccacactgcaccatttgccttagttccctgctgatgcctggcctctt
 cttctttgtcttagACCAGTGAAGAGGTGGTTTACAGAAGATGACTGGACTCAAAGTA
CCCCGTCTCATTTCCCGCAGTAATGACACCCCTTTATATCCCAGAAATGGGAAGGTAG
AGCCCCAGACTCTGTGCGACTATCGAAAGAAAGGATATGTTACTCCTGTCAAAAATC
 AGgt actctccttt cttctgggtg tgcatatgta atctggca
 tgaccttttc ctttttctgc tgctttgttc ttgaggtgaa agggcaccag
 gaaaagaggg caaggaatta aggtacatct ccccatctcc attctgttat
 ttaacctcat ttgtttctgt acatttgggt tgtttctggt ttttctttt
 cttttccctt tttttttttt tttttttttt gagatagagt ctactctgt
 cgtccaggat ggagtgcagt ggtgcaatct tggctcactg caacctacac
 ctcccggtt caagcgattc tctgcctca gcctcctgag tagctgagat

FIGURE 1B

tacaggcacy cgccactacy cctggctaata ttttctatatt ttatagagat
 gcgtttttcac catgttggcc aggctgggtct tgaactgacc tcagggtgac
 cacctgcctc agcctccccaa agtgctggga ttagagtcac gagccatcgc
 ggcctgggttt ttctttatta caaatagtgt tgcaataagc acccttctgc
 atatgttttt gtgcacatgt acaaatatatt atgcaaaaata agtccataaaa
 ttggaattgt taggtcacaa ataactccttt cccccccccc aaattttttt
 tttttttttg agacagcgtc tctgtcacc aggctggagt ccagtggcgc
 aatcatgggt cactgcagcc tcaacgtctc aggctcaagt gattctccaa
 cctcagcctc cctagtagct gggaattaga agcacatgcc accacaccca
 gctaatttta aaaaattttt tgtttagagac aggggttttgc catgctaccc
 aagctgggtct caaatctctg ggctcaagca atctgcccgc ttcggcctcc
 caaagtgtta ggattacaga catgagccac catgcccagc ccaaaaaagt
 ttttgcaatc ttacattctt actagcatga gaatgtcagt tttttcacia
 cccaaacaac acaggattgt atcagcaaga taaacaattg atttaacgtt
 catttaacaa acactttttg acccccagaa cctaccagat gcagtgttag
 gcagcagaga ctcaagatga ctaagacaca acctgtgtcc tcaggaaatc
 tcaatctaaa aaaatagaac aggaaagaaa gaaaaatcta caatctagct
 gcacaaacaa taatagctaa tactttttga gattttattg tttgtcagga
 acttcttaac tctttacatg agtttaata ttttaacct tataacaata
 ttttatgcat agagaaactg agacacaggc aaatttagta acttaccggg
 ggtcacatag ctactgggtg gcaaagtcag ggtagctcc caggacaaat
 gcctccacag ctggtactgt gctctgcttt actgtagcta atagtaaaaa
 tggtagcaaa aatcaatagc agtagaacag tgcaacagat attaaagcga
 agaggaaagac tcacaacaat gacaacattt gtgctgaaat ttttaagaac
 acatggaatt tccttcagcc gggtagagag aagatataga aatgtaaaca
 ccaaagattc atagtctctc tgtatccctt tcagGGTCAGTGTGGTTCCTG
 TTGGGCTTTTAGCTCTGTGGGTGCCCTGGAGGGCCAACTCAAGAAGAAACTGG
 CAAACTCTTAAATCTGAGTCCCCAGAACCTAGTGGATTGTCTGTCTGAGAATGA

FIGURE 1C

TGGCTGTGGAGGGGGCTACATGACCAATGCCTTCCAATATGTGCAGAAGAACCGGG
GTATTGACTCTGAAGATGCCTACCCATATGTGGGACAGgtg agattgctcc
 acacaattat acagctctgt tggctcctcc ttccccagca tgatgtt
 ttgtactgga aacaattcca gaaatactgt tttctgttat cctatcctgc
 tttcttgatg gaataatttc ccacagaagg ccaagaagat ttccacaatc
 tgggggaatt tagggagctt aagctactat agctcctatt tgcatctctg
 ccatggagag aaaacagagg ctaggctacc taccocatag acttccgagc
 tgggttctat aaccctctgc tcaattcctc actcccacaa caaaccacaa
 aaccaccat gctattttca caaattgtgt ggctttatatt tatatgatct
 cagtgtgagt tttcagaaca tttcagcaaa ttatgtaagt ttacatgcta
 acatctataa aatgagagaa aaaacaagtt gcttcatata agagataagg
 gattaactca gttcctcctg catgatcctc tagtcatagg aaggaaatca
 tatctgaaag ggaggcaacc tgaggggttt tttatacaca tagggctggg
 tctgatagac aatataatgt agggccttca caacagaaac ctctgaaaca
 gggacagcaa gtttgagaat aaaaatgatg gctactgtgt tctaagccgt
 gtccttagtg cattttttct tttcttttct ttcatttaat ctcataacaa
 ctctgttagg tagacttctc ttgaatgtat aggtgaggaa atggacactt
 aaggagataa gacagtataa ttcataccac tagtatgtaa caatgtaaga
 tgtatctacc agggatgttt atcttctgca aacattccta ggtatatctc
 ccatgcacat gtgcaagaat ttcttactag gatataatgc cttggaactg
 aattgtctgg gtcttagggg atgtctgtct tcaactttac tacacaatgt
 caaattgttt gccaaaatat ttggaaaaat ttatacctgc aatgtgtaag
 aatccccctt caatcacctt tttatcagta tgtttatctg gccatttgca
 tttcttcttc agtgaattaa ctgtttttat ctcttgctca tttgttttct
 tttttatatt tttgaaatag ggtcttactc tgttgcccaa ggctggagtg
 tggatgataca gtcatagtct actgcagcct ccacttccgg gctcaagcaa
 tcctctcgcc tcagcctccc aaatagctag gatatagggt catgccatca
 tgcccaccaa tttcaaaaaa cctttgaaat tttttttttg taaaagctag
 gcatgggtggc tcatgcctgt aatcccagca ctttgggagg ctgaggtggg

FIGURE 1D

aggatcgctt gagccagga attggaggtc ggcctgatac aacatagcaa
gacctcatct ctacagaaaa aatTTTTTaaa agtagccagg tatgatggcg
tgcatagttc tagctactcc ggaagctggg tgggaggaca acttgagcct
gggagttcaa ggctgctgtg aactgtgac atgtcactgc tctctagcct
gggtgacaga gtgagacct gtCCCCaaaa acaacaaccg TTTTTTTTgg
tagagacatt gtctcgctat gttgccaagg ctagtctcaa actcctgggc
tcaagcaatc ctcccacctc ccaaagtgtc gggattatag atgtaagcca
ccatgcctgg cctacccttt TTTTTTTTTT ttgaaatgga agttttgctt
ttgtcaccta ggcttgagtg cagtggcgcg atcttggtc actgcaacct
ccacctcttg gattcaagca attctctctgc ctccagcctcc tgaatagctg
ggattatagg caccgcgaac cagccccggc tagtttttgt attttttagta
cagacagggg ttcacatgt tggccagctg gtcttgaacc cctgacctca
gggtggctcgc ccgcctcggc ctcccaaaat gctgggatta aaagtgtgag
ccaccatgcc ccaccctta ctcatTTTTa attggattgt TTTTctctt
tcttagcgat tcttaaaagt ttaaagagaa tatttgata caatactatg
tatttaaaag ttgaggtctg tctttccatt ctctcacga tgtcttcaa
tctagaaaag ttaattttta taggcctggc gcggtggctc acgcttgtaa
tcccagcact ttgggaggct gagatgggtg gatcacaagg tcaggagatg
aagaccatcc tggctaacat gggtgaaacc ctgtttctac taaaaaataca
aaaaaattag ctgggcgtgg tggcaggtgc ctgtagtctc agctactcgg
gaggctgagg caggagaatg gcgtgaaccc gggagggtgga gcttgagtg
agccgagatt gcaccactgc actccagcct gggcaactga gcaagactgc
gtttcaaaaa aaaaaaaagt taattttta atagtataat tagtaaaagg
attaattttc cttttgcaat ttttgtaatg tgttttattc gtttatgaat
ggagaaaagg aagaaaaaat aaaattttaa aaagaagaga tgtggccagg
tacggtggct cacacctata atcccagtag tttgggaggc tgaggcaggc
agatcacttg aggtcaggag tttgagacca gctgggataa catggtgaaa
ccccatctct actaaaaata caaaaattag ccagggtgtga ttgcccagc
ttgtaatccc agcaggctga ggcaggagaa ttgctcgaac tcaggaggca

FIGURE 1E

```

gaggttgtag tgagccaaga tcatgccatt gcactccagc ctgggtaaca
gagactctgt ttcaaaaaaa taaaaagata aaaaaggaag agatctgata
gggcygycag ataaacattt taaaggggat ggtattataa gtttgttccc
agcataatgc cagggttattc tgactttaaa gtatcatcac ataatatctt
tttgagtcaa tttccaagat attctgtttc acttgtaatt ctgtgtaatt
tttggcacca ggaggcatca gggatttgga gcacatggca gaaacaaagg
catcttgaaa aatatcaagg cagtagacca ctgtaatctt aaaaaggcat
atcaaatgct gctattgctg ttaatattta gataatgtta gataatgtat
tttttttagag ggtatctcac tatcttgac aggctggagt agagtggcta
ttcacagcat gatcacagta cactaaaggc tcaaactcct gggcacaaac
aatcctcctg cctcagcctg ctgagtagta gataataagt tcttgtggat
gcaaccttag ggttctgaag gggtagtctg taggaaaatg aattgctgaa
agaatacac cacccttaaca tgggctatta ttcgattcca taattgtggc
tgccaatga aacattgcta actacctgta aaatatagtg ttggaagtca
taggctaaat tgctaagttc tttaatctat tttagtgtct tgttatgtac
ttttatattt tgtctttgat gagagcacia ggatcacacc agtccccctg
atataggtgc agagggccca ggtcttcctt cttagctaagc cttggccttg
gcctcctacc cacacagcag ctgggtgcctt cctgccccct gaggctaata
catactatgt ggccagaaga tggtttatgc tttttaaaaa aatcttattt
cagaaatctt tccctactgt tttcctccca catttatgtc ttaaacacc
tgtaggggat tttttttttt tttttttttt tgagatggag tctcgtcttc
gccaggctg gagtgaatg gcgcgatctt ggctcactgc aaggctctgcc
tcccagggtc acgccattct cctgcctcag cctccccagt agctgggact
acaggcgccc gctaccacgc ctggctaatt tttttgcatt tttagtagag
acagggtttc actgttagcc aggatgggtc cgatctcctg acctcgtgat
ccaccctcct cagcctccaa agtgctggga ttaacaggca tggagcccca
ccgcactggc ctgtatttgt gaggaagaac agaccctctt tagaagccct
agactgctgc ctctgttagt. tcaatggcat cactcaaat attggttgag
tttcttactc actgagttgg tttttatgtg tgggtggaagg cgggaatcct

```

FIGURE 1F

cttttcctatc tcgtttctcat tgcctattgc tttgtcctag tcttattaca
 atcttgttttctccagGAAGAGAGTTGTATGTACAACCCACAGGCCAAGGCAGCTA
 AATGCAGAGGGGTACAGAGAGATCCCCGAGGGGAATGAGAAAGCCCTGAAGAGGGCA
 GTGGCCCGAGTGGGACCTGTCTCTGTGGCCATTGATGCAAGCCTGACCTC
 CTTCAGTTTTACAGCAAAG gtaagaagct gctgatccta tacagcactg
 tcttttatga tacaaacttg atggtttctc gaaggacctt ggggtattttc
 agtacttagt ttttgtattc acatggaggt ggccagagag aaatttaacaa
 ctgctgcagt atggagcagc atctctgtgg taaacctcc tgacacggat
 ggaattcttc aaacagtctc ctagactggg agatcccaca ggggtgacctt
 tggattgcat agagcctcac gctggtagtt tgtattctag GTGTGTATTA
 TGATGAAAGCTGCAATAGCGATAATCTGAACCATGCGGTTTTGGCAGTGGGAT
 ATGGAATCCA GAAGGGAAACAAGCACTGGATAATTAAAAACAGgtaatgatg
 ggaacactac ttttgttatt cagtcacctt ttttaacactc aacctcacct
 ccagcttccc gatattcctt tctctgtccc aaatcaagaa aaaattatct
 cagagttctc acttctatct tctcagtcag aggtctctta tctcagttct
 gacacttaat ggccagtggt ttagtccatt ttgcattgcc aaaaaagaat
 acccgagaft gggtagttta taaagaaacg aggtttgttt ggctatacaa
 agcgtggcac tagtatctgc tcagcctctg atgaggcctc agagctttta
 ctcatggcag aaggcaaaag agggagcagg catgtcacat agtgagagag
 ggagcaagag agagagggag gtgcccactc tttaaagaac cagctcttgc
 atgaactaat agagtgagaa ctcactcacc accaaggcga tggcaccaag
 ccattccatg aggaatccac tctcataacc caaacacctc ccactatgcc
 ccacctcca cattggggat cacatctcag catgagactg ggaggggaca
 cacatccaaa ccatatccgc cagacaatag tgctcaatta tgtgctgggc
 agatgctccc tgtgtgcaag gtgcttagtg acatacataa accaacgagc
 agatgacacc ttcagtgcgc tcagagccca ataagacaga cctaactaac
 catgagataa agcagtacaa agaaccagca ggagcttttg aattacgtat
 ttttactttc ttttgtctct aatgtgatca gtttcttaga tggtttccat
 tagcaatctg tctttaacag taggggagca gcgttaaagg tttaatatc

FIGURE 1G

```

cttttgaaca gtttttttcc ttcaaaatac acttaagata cacgtatata
agaacttgcc aaagattgtg aagagaaaca ttttttagaa ataagatata
aacaaaaaaa gttagtgtta ctttcctatg ttggggaaca aagaaaactc
cagggtacct tgcttcccat ttctcttttag cacttggtga cttttgggga
ggggcagatt gataacaatt atagttttcc tttcctggct gatcaccatt
aacctggcag cagcactggc taaatctcct gtcccttagtg ccctccaagg
agcaggagcc ctagactctg ggtcgtgac agactcacgc agtgggtgtg
ttcaaacctg aagcaacttt ttatatcaca gttccaactc aagggtgaacc
tgagcatctt cccaagtctc ccacagcttc tgtcctgtgt tgtcccttct
cttgactccc aggtccaagc acttaccctg ttctttcatg atcagggtacc
atgtgtggag atagcttcca agagagctgg gaggaagaaa ggacacaccc
gggcaggatc aggaacactg gggggccctg gagaagggga gagtggggga
gggtacaggt tttaaataaa atgtgttggg aattagagaa ttgctggttg
gggaaagagg tctgaaaaa attcaggaag ataaacaaga caatctctcc
tctctcctct ttctcacgtc gtctctcttg tcttctagtc tgcgtactca
tttccttagt aatctcatcc actctcatag ttcatccat ctctcctatg
gggtttacc ccaaatcaag atcaccagct tcagcctcct tcttatgctc
taaactcaca ttttcaagat taatattccc caaatacagc tctgatcata
tcactctccc actcaaaatc cctcactggc tcctcacgat gatgggtcac
agagtaaagg tgaagctttt taaccttgca gtaaaggtaa ttcaacctga
tctcaatctg cttttccaga catctctccc actacaccct gttaggcaca
ctgcttttca gctacatgat cctaacagtg cccacactt tctgcctct
gttgttcatt tcacaccctt ccactggcat ccccttccca caggctgaaa
ttctacttag ctttttggct cagctcaaat gccacctctt acatcaagcc
tctaagattc tcttgatcag aaggaatctt tccctccttt gatacctaca
gtattatgcc ttctccctat ttcttgactt taaactcttt aaagttaaaa
aacatcatat tcatttttgt gtaccatcag tacctcgcac aatactcagt
aaatatttta atgaataaat aaactgagag tactaagtat ttttcttgat
tggtcttacag CTGGGGAGAAAACCTGGGGAACAAAGGATATATCCTCATGGCT

```


FIGURE 1H

CGAAATAAGAACAAACGCCTGTGGCATTGCCAACCTGGCCAGCTTCCCCAAGATGTG
ACTCCAGCCAGCCCAAATCCATCCTGCTCTTCCATTTCCTTCCACGATGGTGCAGT
GTAACGATGCACCTTTGGAAGGGAGTTGGTGTGCTATTTTTGAAGCAGATG
TGGTGATACTGAGATTGTCTGTTTCAGTTTCCCCATTTGTTTGTGCTTCAAAATGA
TCCTTCTTACTTTTGCTTCTCTCCACCCATGACCTTTTTCCTACTGTGGCCATCAGGA
CTTTCCTGACAGCTGTGTACTCTTAGGCTAAGAGATGTGACTACAGCCTGCCCC
TGACTGTGTGTTGTCCTCCAGGGCTGATGCTGTACAGGTACAGGCTGGAGATTTTCACAT
AGGTTAGATTCTCATTACAGGGACTAGTTAGCTTTAAGCACCTTAGAGGACTAGGG
TAATCTGACTTCTCACTTCCTAAGTTCCCTTCTATATCCTCAAGGTAGAAATGTCT
ATGTTTTCTACTCCAATTCAATAATCTATTCAATAAGTCTTGGTACAAGTTTACAT
GATAAAAAGAAATGTGATTTGTCTTCCCTTCTTTGCACCTTTTGAAATAAAGTATTT
ATCTCCTGTCTACAGTTTAATAAATAGCATCTAGTACACATTCATTTTGTGTTGGA
TACTGTGTTAGGTGCTGGAGGAAAAAAGATGAATAGAACATCTTCTATGTACTTGA
TGCGCTCACAGTCTGGTTGTAGAGACTGTCACATAAACATTTTCATCCCAATTCATT
TATTTGTTTCATTCCTTCAGCCAATATATATTGAGTTCTTACTCTGTGCCAAGAACT
GTA
GTACTACATTTCTGGGATTAAGTGGATATAAG
GAGATCTCAGTGTTTAATCTGCCTGAGGGGAGACTAAATTAAGTGACATGGAACT
TGGGTCTTGAAAAACATTTTAAGGTTATTTTTCTTTTCTCTCTCTCTCGCT
CTGTCTTTCTC TCTCTTTCGTGAGGCTCCTCTGTTGCCAGGCTGGAGTC
AGTGGCACTCATAGCTCACTGCAGCCTTGATCTCCTGGGCTCAAGAGTTCTTCCCA
CCTCAGTCTCCTAAGTAGCTTGACTACGG

FIGURE 2

cDNA CACACTTTGCTGCCGAAACGAGCCGACACAACAGATTTCATCAGAG¹G 49
 intron 1
 ATGTGGGGGCTCAAGTTCTGCTGCTACCTGTGGTGAGCTTTGCTCTGTACCCCTGAGGAG 109
 M W G L K V L L L P V V S F A L Y P E E
 intron 2
 ATACTGGACACCCACTGGGAGCTATGGAAGAAGACCCACAGGAAGCAATATAACAACAAG¹ 169
 I L D T H W E L W K K T H R K Q Y N N K
 GTGGATGAAATCTCTCGGCGTTTAATTGGGAAAAAACCTGAAGTATATTTCCATCCAT 229
 V D E I S R R L I W E K N L K Y I S I H
 AACCTTGAGGCTTCTCTGCTGCTCCATACATATGAACTGGCTATGAACCACTGGGGGAC 289
 N L E A S L G V H T Y E L A M N H L G D
 intron 3
 ATG¹ACCAGTGAAGAGGTGGTTCAGAAGATGACTGGACTCAAAGTACCCCTGTCTCATTC 349
 M T S E E V V Q K M T G L K V P L S H S
 CGCAGTAATGACACCCCTTTATATCCAGAAATGGGAGGTAGAGCCCCGAGACTCTGTCCGAC 409
 R S N D T L Y I P E W E G R A P D S V D
 intron 4
 TATCGAAGAAAGGATATGTTACTCCTGTCAAAAATCAG¹GGTCAGTGTGGTTCCTGTGG 469
 Y R K K G Y V T P V K N Q G Q C G S C W
 GCTTTAGCTCTGTGGGTGCCCTGGAGGGCCAACTCAAGAAGAAAAGTGGCAAACTCTTA 529
 A P S S V G A L E G Q L K K K T G K L L
 AATCTGAGTCCCCAGAACCTAGTGGATTGTGTGCTGAGAATGATGGCTGTGGAGGGGGC 589
 N L S P Q N L V D C V S E N D G C G G G
 TACATGACCAATGCCTTCCAATATGTGCAGAAGAACCGGGGTATTGACTCTGAAGATGCC 649
 Y M T N A F Q Y V Q K N R G I D S E D A
 intron 5
 TACCCATATGTGGGACAG¹GAAGAGAGTTGTATGTACAACCCAACAGGCAAGGCAGCTAAA 709
 Y P Y V G Q E E S C M Y N P T G K A A K
 TGCAGAGGGTACAGAGAGATCCCCGAGGGGAATGAGAAAGCCCTGAAGAGGGCAGTGGCC 769
 C R G Y R E I P E G N E K A L K R A V A
 CGAGTGGGACCTGTCTCTGTGCGCCATTGATGCAAGCCTGACCTCCTCCAGTTTTACAGC 829
 R V G P V S V A I D A S L T S F Q P Y S
 intron 6
 AAAG¹GTGTGTTATGATGAAAGCTGCAATAGCGATAATCTGAACCATGCGGTTTTGGCA 889
 K G V Y Y D E S C N S D N L N H A V L A
 intron 7

FIGURE 2A

GTGGGATATGGAATCCAGAAGGGAACAAGCACTGGATAATTAAAAACAG¹CTGGGGAGAA 949
V G Y G I Q K G N K H W I I K N S W G E

AACTGGGGAAACAAAGGATATATCCTCATCGCTCGAAATAAGAACAACGCCTGTGGCATT 1009
N W G N K G Y I L M A R N K N N A C G I
7R←—————8P

GCCAACCTGGCCAGCTTCCCCAAGATGTGACTCCAGCCAGCCCAATCCATCCTGCTCTT 1069
A N L A S F P K M

CCATTTCCTTCCACGATGGTGCAGTGTAAAGATGCACCTTTGGAAGGCAGTTGGTGTGCTA 1129

TTTTTGAAGCAGATGTGGTGATACTGAGATTGTCTGTTCAGTTTCCCCATTGTGTGTGC 1189

TTCAAATGATCCTTCTACTTTGCTTCTCTCCACCCATGACCTTTTCCACTGTGGCCAT 1249
8R←—————

CAGGACTTTCCTTCCAGCTGTGTACTCTTAGGCTAAGAGATGTGACTACAGCCTGCCCT 1309
—————9P

CTGACTGTGTGTGCCAGGGCTGATGCTGTACAGGTACAGGCTGGAGATTTTCACATAGG 1369

TTAGATTCTCATTACAGGGACTAGTTAGCTTTRAGCACTTAGAGGACTAGGGTAATCTG 1429

ACTTCTCACTTCCCTAAGTTCCCTTCTAATCTCTCAAGGTAGAAATGTCTATGTTTTCTAC 1489

TCCATTTCATAAATCTATTTCATAAGTCTTTGGTACAGTTTTCATGATAAAAGAAATGT 1549

GATTTGTCTTCCCTTCTTTGCACTTTTGAATAAAGTATTTATCTCTCTCTACAGTTTA 1609
9R←—————

ATAAATAGCATCTAGTACACATTCA 1634

3' UTR TTTTGTGTGGATACTGTGTAGGTGCTGGAGGAAAAAGATGAATAGAACATC 1688

TTCTATGTACTTGATGCGCTCACAGTCTGGTTGTAGAGACTGTCACATAAACATTTTCATC 1748

CCAATTCAATTTATTTGTTCAATTCCTTCAGCCAATATATATTGAGTTCTTACTCTGTGCCA 1808

AGAACTGTACTACATTTCTGGGATTAAGTGGATATAAGGAGATCTCAGTGTTTAATCTGC 1868

CTGAGGGGAGACTAAATTAAGTGACATGGAACTTGGGTCTTGAAAAACATTTTAAGGTT 1928

ATTTTTTCTTTTCTCTCTCTCGCTCTGTCTTTCTCTCTCTTTCCTCAGGGTCTCCCTC 1988

TGTTGCCAGGCTGGAGTCACTGGCACTCATAGCTCACTGCAGCCTTGATCTCCTGGGCT 2048

CAAGAGTTCTTCCACCTCAGTCTCCTAAGTAGCTTGGACTACGG 2108

FIGURE 3

Cathepsin-K Sequence: Exon-Intron Boundaries

(A) 5' Untranslated sequence [SEQ ID NO. 2]

```

GCTTTGGCTC CCAAAGGCCT GGGATTACAG GCGTGAACCA CTGCGCCTAG
CCTGTTAGCA GCTCTTAAAA TCCAGAGGCA TAAGCCTGTA TTTTGTAGGG
TTTATGCATG GAATCCAGCT AGAAACTGAG TCTATTACAG ATCCCATTTA
TTATCCTTTC TATTCCAAGA AGCCTTTTTT TCTCCTTCCC CACATCTGTT
TATGGAAGAA AATGAAGTTT GGGGTGTGGT TTGAGGAATC AGCTAGATTC
TTATGATCTG TCACATGCTT GGATGTTGGG GAAGCATTTG GAGAAGCTCA
TGTGACTTGT CCTAGATTGG GGATTTTAAAT TGAGACAGAT GATGTTTATC
GGGCATCCCA CCACCTGAGA GTTTTAGCAA CAGAGTCACA TGTGAGTCCA
TCAGAACTTA CGGCATTGAT TCAAGTGCTG TCATAAATAA CCAGGACTGC
TGTTTTTGGT TACTTTTAAA GACAGTTTCA TCTGGACTTT CTGGGCATAT
CCTCCTTCAG CAAAACCACA TTAGGCTGGG AAAACTATTC TGCCTGGAAG
TAATGACAAC TTGCAACCAA CAAGCTTATA AAAATACAAA GAATTCTGGA
GCCTATGGCT TCCATTACAT TATTCTTTTA TAGCCTTTTA TGTTTCATTAC
CGCATCCAG AGGTGAGAGT CAGACACAAA TATGAAAATA GGTTCATG
TTGGAGAGGT AAATCCTAAC AGGAAAGGGG TAGGAAAAGA TATAATCCCC
CAATATTAAA ATAAAGATAT TGAAGAAGAA GGATGGGAGA GACTAGGGCT
GTGTCCTTCC TTTTACTCAC CAAAAGAGAA AGTAAGCTCC TATTTGAGTC
AATAGATATT GAGGTCTTGT TATTTGCCAC CAAAGACAGT CTGTGAGAC
TAAATAGCTA GTAATTCCCT ACCCTGGCAC ACATGCTGCA TACACACAGA
AACACTGCAA ATCCACTGCC TCCTTCCCTC CTCCCTACCC TTCCTTCTCT
CAGCATTTCT ATCCCCGCCT CCTCCTCTTA CCCAAATTTT CCAGCCGATC
ACTGGAGCTG ACTTCCGCAA TCCCGATGGA ATAAATCTAG CACCCCTGAT
GGTGTGCC

```

FIGURE 3A

(B) Exon 1 [SEQ ID NO. 3]

CACACTTTGCTGCCGAAACGAAGCCAGACAACAGATTTCCATCAGCAG

(C) Intron 1 [SEQ ID NO. 4]

gtaacgtttg caacttccta gatcttttag cttttcattc ctgtcaattc
tctgagtatt agggatgtag tgacttgagg atcacaataa acttttagcc
tctgcagatg aaaacagaga tgcacttctt aggtcattcc ctggctaaat
aaaatctgcc tggaaatctg tagaattcct tgtatgattt atatataac
atacatgatt gttagtaaaa gcaaagtata tagggaatca tttccccatc
cttcaagagt ggcctttctg cagtgttttc tactttggcc aacaaggatc
aaaacgggta actccttagt gaggaggagg agagtgggtat ggggaggtag
tagctcagtg cttcctgttc actgagacat ctcaaagccc ttaacactct
agttttttaa tgtcctactg gacatttttg cagtttgcaa aattacatgt
aaatggacta taagcaattg tgtaagccat atgtcatgct gcaggctgca
aattgttctt aaaatggagg atttgtaact aagaagcca atgcaagaaa
tgagtgaagc taactagagt aaacttatga aaagctgtga atttcatcat
catagaacat tgcttttcag tctgaacatt cttctaacaa accttggatc
tgaggcttct tgtccttttg ggcagccaca gtgggttttt gttgttaggg
gaaaataaaa aaccttgccc gcagcatctg gtttaagatta gggcagtttc
ctgcctaagg aggggaaggga gagaanaagg aagaagaaat gcataaggag
aatgaggaga tatacaatgt ctcagaaaac aggaacatt gtcctatttt
cccttgctct cttctgacaa gatctgggaa agtaccagaa tttaggcacg
aaagagaaga acgcctcgaa gaaatgatca ggaagcaaaa cttagacgga
aatctctcct ttgtgtattc tgaacccac taccaccttg ctatttgtct
gtctccaagc ctgctaggga ccctggagga aacgcactga gccattctg
attgtccagt ttctatcccc cttttctggg tgtgtacgtg tgtgtgtgtg
tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgagagagag agagacagag
agagaaaacag agagagtgtg tgttgccctaa atctcccag agagagagag
agagagagag agagagagag agagagaaaa gagagaaatg gctaaatccc
cctagatcaa agtccttgga accagatgta ccagcatcct atctaaacac
aggccccctc tgactatcat tgttttatca ccctttttcc gtctaccttt
ctcttctca taaagcctag ttttctcttg tttccctgcc aaatggaaga
gttttcccta actacattct tctgcag

FIGURE 3B

(D) Exon 2 [SEQ ID NO. 5]

GATGTGGGGGCTCAAGGTTCTGCTGCTACCTGTGGTGAGCTTTGCTCTGTA
CCCTGAGGAGATACTGGACACCCACTGGGAGCTATGGAAGAAGACCCACA
GGAAGCAATATAACAACAAG

(E) Intron 2 [SEQ ID NO. 6]

gtgcctgggg tcctggaggg ggcattggcag gaaggctgag acctgagctc
tctcatctta gcttccagac tcccttcttc aatccaaatg ctttattcca
agcaaatcag tccctcttcc ctaactcatg ttaacatacg gttttcatte
ctatgcttca atcatcctct tgtcaaaact gtattccttc cctttggttt
tataagtgtg taacattcct cttttgggaa gagtcccaag attaatgctg
ttaatccata agcaattttt ctgtctctcc agagcttgtg tggttgttta
catattatct ctcttcttgc aggctcttaa ttccatgggt agttccccaa
ctaaactgta aacttttatg attgtgagtt tccctttattc tccataaacc
cttcacaafa ttacatatga actgtagaca gtctatacaa gtactgacta
tgctttgttt ag

(F) Exon 3 [SEQ ID NO. 7]

GTGGATGAAATCTCTCGGCGTTTAATTTGGGAAAAAACCTGAAGTATATTTCCAT
CCATAACCTTGAGGCTTCTCTGGTGTCCATACATATGAACTGGCTATGAACCACC
TGGGGGACATG

(G) Intron 3 [SEQ ID NO. 8]

gcaagtatag cttcagctcc tgtcccacct gcaccatttg ctttagttcc
ctgctgatgc ctggcctctt tcttctttgt ctttag

FIGURE 3C

(H) Exon 4 [SEQ ID NO. 9]

ACCAGTGAAGAGGTGGTTCAGAAGATGACTGGACTCAAAGTACCCCTGTCTCATTC
CCGCAGTAATGACACCCTTTATATCCCAGAATGGCAAGGTAGAGCCCCAGACTCTG
TCGACTATCGAAAGAAAGGATATGTTACTCCTGTCAAAATCAG

(I) Intron 4 [SEQ ID NO. 10]

gtactctcct ttcttctggg tgtgcatatg taatctggca tgaccttttc
ctttttctgc tgctttgttc ttgaggtgaa agggcaccag gaaaagaggg
caaggaatta aggtacatct cccattccc attctgttat ttaacctcat
ttgtttctgt acatttgggt tgtttctgggt ttttcttttt cttttccctt
tttttttttt tttttttttt gagatagagt ctactctgt cgtccaggat
ggagtgcagt ggtgcaatct tggctcactg caacctacac ctcccggtt
caagcgattc tcctgcctca gcctcctgag tagctgagat tacaggcacg
cgccactacg cctggctaatt ttttctattt ttatagagat gcgttttcac
catgttgggc aggctggtct tgaactgacc tcagggtgac cacctgcctc
agcctcccaa agtgctggga ttagagtcac gagccatcgc ggcctggttt
ttctttatta caaatagtgt tgcaataagc acccttgtgc atatgtttt
gtgcacatgt acaaatattt atgcaaaata agtcctaata ttggaattgt
taggtcacaataaatccttt cccccccccc aaattttttt tttttttttg
agacagcgtc tctgtcacc caggctggagt ccagtggcgc aatcatggct
cactgcagcc tcaacgtctc aggtcgaagt gattctccaa cctcagcctc
cctagtagct ggaattaga agcacatgcc accacaccca gctaatttta
aaaaattttt tgtagagac aggtttttgc catgctaccc aagctggtct
caaatcctg ggctcaagca atctgcccgc ttcggcctcc caaagtgcta
ggattacaga catgagccac catgcccagc ccaaaaaagt ttttgcaatc
ttacattctt actagcatga gaatgtcagt tttttcaca cccaaacaac
acaggattgt atcagcaaga taaacaattg atttaacgtt catttaacaa
acactttttg acccccagaa cctaccagat gcagtgttag gcagcagaga

FIGURE 3D

ctcaagatga ctaagacaca acctgtgtcc tcaggaaatc tcaatctaaa
 aaatagaac aggaagaaa gaaaaatcta caatctagct gcacaaacaa
 taatagctaa tactttttga gattttattg tttgtcagga acttcttaac
 tctttacatg agttttaata tttaatccct tataacaata ttttatgcat
 agagaaactg agacacaggc aaatttagta acttaccggg ggtcacatag
 ctactgggtg gcaaagtcag ggtagctcc caggacaaat gcctccacag
 ctgggtactgt gctctgcttt actgtagcta atagtaaaaa tggtagcaaa
 aatcaatagc agtagaacag tgcaacagat attaagcggg agaggaagac
 tcacaacaat gacaacattt gtgctgaaat ttttaagaac acatggaatt
 tccttcagcc gggtagagag aagatataga aatgtaaaca ccaaagattc
 atagtttctc tgtatccctt tcag

(J) Exon 5 [SEQ ID NO. 11]

GGTCAGTGTGGTTCCTGTTGGGCTTTTAGCTCTGTGGGTGCCCTGGAGGGCCAA
 CTCAGAAGAAAACCTGGCAAACCTCTTAAATCTGAGTCCCCAGAACCTAGTGGATTG
 TGTGCTCGAATGATGGCTGTGGAGGGGGCTACATGACCAATGCCTTCCAATATG
 TGCAGAGAACCGGGGTATTGACTCTGAAGATGCCTACCCATATGTCCGACAG

(K) Intron 5 [SEQ ID NO. 12]

gtgagattgc tccacacaat tatacagctc tgttggctcc tccttcccca
 gcatgatgtt ttgtactgga aacaattcca gaaatactgt tttctgttat
 cctatcctgc tttcttgatg gaataatttc ccacagaagg ccaagaagat
 tccacaaatc tgggggaatt tagggagctt aagctactat agctcctatt
 tgcattcttg ccatggagag aaaacagagg ctaggctacc taccocatag
 acttccgagc tgggttctat aacctctgc tcaattcctc actccacaa
 caaaccaca aaccaccat gctattttca caaatttgtt ggcttttatt
 tatatgatct cagtgtgagt tttcagaaca tttcagcaa ttatgtaagt
 ttacatgcta acatctataa atgagagaa aaaacaagtt gcttcatata

FIGURE 3E

agagataagg gattaactca gttcctcctg catgatccctc tagtcatagg
aaggaaatca tatctgaaag ggaggcaacc tgagggggttt ttatatacaca
tagggctggg tctgatagac aatataatgt agggccttca caacagaaac
ctctgaaaca gggacagcaa gtttgagaat aaaaatgatg gctactgtgt
tctaagccgt gtccttagtg catTTTTTct tttctTTTT ttcatttaat
ctcataacaa ctctgttagg tagacttctc ttgaatgtat aggtgaggaa
atggacactt aaggagataa gacagtataa ttcataccac tagtatgtaa
caatgtaaga tgtatctacc agggatgttt atcttctgca aacattccta
ggtatatctc ccatgcacat gtgcaagaat ttcttactag gatataatgc
cttggaactg aattgtctgg gtcttagggg atgtctgtct tcaactttac
tacacaatgt caaattgttt gccaaaatat ttggaaaaat ttatacctgc
aatgtgtaag aaatccccctt caatcacctt tttatcagta tgtttatctg
gccatttgca tttcttcttc agtgaattaa ctgtttttat ctcttgctca
tttgtttttc tttttatttt ttgaaatag ggtcttactc tgttgcccaa
ggctggagtg tggatgatac gtcatagctc actgcagcct ccacttccgg
gctcaagcaa tctctctgcc tcagcctccc aaatagctag gatataagggtg
catgccatga tggccaccaa tttcaaaaaa cctttgaaat tttttttttg
taaaagctag gcatgggtggc tcatgcctgt aatcccagca ctttgggagg
ctgaggtggg aggatcgctt gagcccagga attggagggtc ggcctgatac
aacatagcaa gacctcatct ctacagaaaa aatttttaaa agtagccagg
tatgatggcg tgcatagttc tagctactcc ggaagctggg tgggaggaca
acttgagcct gggagttcaa ggctgctgtg aactgtgac atgtcactgc
tctctagcct ggggtgacaga gtgagaccct gtcccaaaa acaacaaccg
tttttttttg tagagacatt gtctcgctat gttgccaagg ctagtctcaa
actcctgggc tcaagcaatc ctccacctc ccaaagtgtc gggattatag
atgtaagcca ccatgcctgg cctacccttt tttttttttt ttgaaatgga
agttttgctt ttgtcaccta ggcttgagtg cagtggcgcg atcttggtc
actgcaacct ccacctcctg gattcaagca attctcctgc ctacgcctcc
tgaatagctg ggattatagg caccgcaac cagccccggc tagtttttgt
attttttagta cagacagggg ttccacatgt tggccagctg gtcttgaacc
cctgacctca ggtgggtccgc ccgcctcggc ctcccaaat gctgggatta

FIGURE 3F

aaagtgtgag ccaccatgcc ccacccctta ctcatttttta attggattgt
tttttctctt tcttagcgat tcttaaaagt ttaaagagaa tatttggata
caatactatg tattttaaag ttgaggtctg tctttccatt cttctcacga
tgtctttcaa tctagaaaag ttaattttta taggcctggc gcggtggctc
acgcttgtaa tcccagcact ttgggaggct gagatgggtg gatcacaagg
tcaggagatg aagaccatcc tggctaacat gggtgaaacc ctgtttctac
taaaaataca aaaaaattag ctgggcgtgy tggcagggtg ctgtagtccc
agctactcgy gaggctgagg caggagaatg gcgtgaaccc gggagggtga
gcttgacgtg agccgagatt gcaccactgc actccagcct gggcaactga
gcaagactgc gtttcaaaaa aaaaaaaagt taattttta atagtaaaat
tagtaaaagg atteattttc cctttgcaat ttttgtaatg tgttttattc
gtttatgaat ggagaaagg aagaaaaaat aaaattttaa aaagaagaga
tgtggccagg tacggtggct cacacctata atcccagtag tttgggaggc
tgaggcaggc agatcacttg aggtcaggag tttgagacca gctgggataa
catggtgaaa ccccatctct actaaaaata caaaaattag ccagggtgtga
ttgcgcacgc ttgtaatccc agcaggctga ggcaggagaa ttgctcgaac
tcaggaggca gaggttgacg tgagccaaga tcatgccatt gcactccagc
ctgggtaaca gagactctgt ttcaaaaaaa taaaagata aaaaaggaag
agatctgata gggcggccag ataaacattt taaaggggat ggtattataa
gtttgttccc agcataatgc caggttattc tgacttttaa gtatcatcac
ataatatctt tttgagtcaa tttccaagat attctgtttc acttgtaatt
ctgtgtaatt tttggcacca ggaggcatca gggattttga gcacatggca
gaaacaaagg catcttgaaa aatatcaagg cagtagacca ctgtaatctt
aaaatggcat atcaaatgct gctattgctg ttaatattta gataatgtta
gataatgtat ttttttagag ggtatctcac tatcttgac aggctggagt
agagtggcta ttcacagcat gatcacagta cactaaaggc tcaaactcct
gggcacaaac aatcctcctg cctcagcctg ctgagtagta gataataagt
tcttggtgat gcaaccttag ggttctgaag gggtagtctg taggaaaatg
aattgctgaa aagaatacac caccttaaca tgggctatta ttcgattcca
taattgtggc ttgccaatga aacattgcta actacctgta aaatatagt
ttggaagtca tagyctaaat tgctaagttc tttaatctat tttagtgtct

FIGURE 3G

tggttatgtac ttttatattt tgtctttgat gagagcacaa ggatcacacc
agttccccctg atataggtgc agagggccca ggtcttccct ctagctaagc
cttggccttg gcctcctacc cacacagcag ctggtgcctt cctgccccct
gaggctaata catactatgt ggccagaaga tgggttatgc tttttaaaaa
aatcttattt cagaaatctt tccctactgt tttcctccca catttatgtc
ttaaaacacc tgtaggggat tttttttttt tttttttttt tgagatggag
tctcgctctc gccaggtcg gagtgcattg gcgcgatctt ggctcactgc
aaggtctgcc tcccagggtc acgccattct cctgcctcag cctccccagt
agctgggact acaggcgccc gctaccaagc ctggctaatt tttttgcatt
tttagtagag acagggtttc actgttagcc aggatgggtc cgatctcctg
acctcgatgat ccaccctcct cagcctccaa agtgctggga ttaacaggca
tggagcccca ccgcactggc ctgtatattt gaggaagaac agaccctctc
tagaagccct agactgctgc ctctgttagt tcaactggcat cactcaaaat
attggttgag tttcttactc actgagttgg tttttatgtg tggtggaagg
cgggaatcct cttttcatac tcgttctcat tgcctattgc tttgtcctag
tctattaca atcttgtttc ttoag

(L) Exon 6 [SEQ ID NO. 13]

GAAGAGAGTTGTATGTACAACCCAACAGGCAAGGCAGCTAAATGCAGAGGGTACAG
AGAGATCCCCGAGGGGAATGAGAAAGCCCTGAAGAGGGCAGTGGCCCCAGTGGGAC
CTGTCTCTGTGGCCATTGATGCAAGCCTGACCTCCTTCCAGTTTACAGCAAAG

(M) Intron 6 [SEQ ID NO. 14]

gtaagaagct gctgatccca tacagcactg tcttttatga tacaacttg
atggtttctc gaaggacctt gggatatttc agtacttagt ttttgtattc
acatggagggt ggccagagag aaattaacaa ctgctgcagt atggagcagc
atctctgttg taaaccctcc tgacacggat ggaattcttc aaacagtctc
ctagactggg agatcccaca gggtgacctt tggattgcat agagcctcac
gctggtagtt tgtattctag

FIGURE 3H

(N) Exon 7 [SEQ ID NO. 15]

GTGTGTATTATGATGAAAGCTGCAATAGCGATAATCTGAACCATGCGGTTTGGCA
GTGGGATATGGAATCCAGAAGGGAACAAGCACTGGATRAATTAATAACAG

(O) Intron 7 [SEQ ID NO. 16]

gtaatgatgg gaacactact tttgttattc agtcaccctt ttaacactca
acctcacctc cagcttcccg atattccttt ctctgtccca aatcaagaaa
aaattatct cagagttctc acttctatct tctcagtcag aggtctcttaa
ttctcagttc gacacttaat ggccagtgtg ttagtcattt ttgcattgcc
acaaaagaat acccgagact gggtagttta taaagaaacg aggtttgttt
ggctatacaa agcgtggcac tagtatctgc tcagcctctg atgaggcctc
agagctttta ctcatggcag aaggcaaaag agggagcagg catgtcacat
agtgaagag ggagcaagag agagagggag gtgccgactc tttaaagaac
cagctcttgc atgaactaat agagtgaag ctcactcacc accaaggcga
tgccaccaag ccattccatg aggaatccac tctcataacc caaacacctc
ccactatgcc ccacctccca cattggggat cacatcttcag catgagactg
ggaggggaca cacatccaaa ccataccgc cagacaatag tgctcaatta
tgtgtgggc agatgctccc tgtgtgcaag gtgcttagtg acatacataa
accaacgagc agatgacacc ttcagtgaac tcagagccca ataagacaga
cctaactaac catgagataa agcagtacaa agaaccagca ggagctttgg
aattacgtat ttttactttc ttttgtctct aatgtgatca gtttcttaga
tggtttccat tagcaatctg tctttaacag taggggagca gcgttaaagg
tttaatatc cttttgaaca gtttttttcc ttcaaaatac acttaagata
cacgtatata agaacttgcc aaagattgtg aagagaaaca ttttttagaa
ataagatata aacaaaaaaa gttagtgtta ctttcctatg ttggggaaca
aagaaaactc cagggtacct tgcttcccat ttctcttttag caccttgtga
cttttgggga ggggcagatt gataacaatt atagttttcc tttcctggct
gatcaccatt aacctggtag cagcactggc taaatctcct gtccttagtg
ccctccaagg agcaggagcc ctagactctg ggtcgctgac agactcacgc

FIGURE 31

agtgggtgttg ttcaaacctg aagcaacttt ttatatcaca gttccaactc
aagggtgaacc tgagcatctt cccaagtctc ccacagcttc tgtcctgtgt
tgtcccttct cttgactccc aggtccaagc acttaccctg ttctttcatg
atcaggtacc atgtgtggag atagcttcca agagagctgg gaggaagaaa
ggacacaccc gggcaggatc aggaacactg ggggccccctg gagaagggga
gagtggggga ggttacaggt tttaaataaa atgtgttggg aattagagaa
ttgctgggttg gggaaagagg tctgaaaaca attcaggaag ataaacaaga
caatctctcc tctctctctt ttctcacgtc gtctctcttg tcttctagtc
tcgctactca tttccttagt aatctcatcc actctcatag tttcatccat
ctctccatg gggtttacc ccaaatcaag atcaccagct tcagcctcct
tcttatgtct taaactcaca ttttcaagat taatattccc caaatacagc
tctgatcata tcaactctcc actcaaaatc cctcaactggc tcttcacgat
gatgggtcac agagtaaagg tgaagctttt taacottgca gtaaaggtaa
ttcaacctga tctcaatctg cctttccaga catctctccc actacacct
gttaggcaca ctgcttttca gctacatgat cctaacagtg cccacacgtt
tcctgcctct gttgttcatt tcacacctt ccaactggcat ccccttccca
caggtcgapa ttctacttag ctttttggct cagctcaaat gccacctctt
acatcaagcc tctaagattc tcttgatcag aaggaatctt tccctccttt
gataacctaca gtattatgcc ttctccctat ttcttgactt taaactcttt
aaagttaaaa aacatcatat tcatttttgt gtaccatcag tacctcgac
aatactcagt aaatatttta atgaataaat aaactgagag tactaagtat
ttttcttgat tggctctaca g

(P) Exon 8 [SEQ ID NO. 17]

CTGGGGAGAAAACCTGGGGAACAAAGGATATATCCTCATGGCTCGAATAAGAACA
ACGCCTGTGGCATTGCCAACCTGGCCAGCTTCCCAAGATG End Coding

FIGURE 3J

(Q) 3' Untranslated sequence cDNA [SEQ ID NO. 18]

TGACTCCAGCCAGCCCAAATCCATCCTGCTCTTCCATTTCCCTCCACGATGGTG
CAGTGTAACGATGCACCTTTGGAAGGGAGTTGGTGTGCTATTTTTGAAGCAGATGTC
GTGATACTGAGATTGTCTGTTCAGTTTCCCCATTTGTTTGTGCTTCAAATGATCCT
TCCTACTTTGCTTCTCTCCACCCATGACCTTTTTCCACTGTGGCCATCAGGACTTT
CCCCTGACAGCTGTGTACTCTTAGGCTAAGAGATGTGACTACAGCCTGCCCTGAC
TGTGTTGTCCCAGGGCTGATGCTGTACAGGTACAGGCTGGAGATTTTCACATAGGT
TAGATTCTCATTACGGGACTAGTTAGCTTTAAGCACCCCTAGAGGACTAGGGTAAT
CTGACTTCTCATTCCCTAAGTTCCCTTCTATATCCTCAAGGTAGAAATGTCTATGT
TTTCTACTCCAATTCATAAATCTATTTCATAAGTCTTTGGTACAAGTTTACATGATA
AAAAGAAATGTGATTTGTCTTCCCTTCTTTGCACCTTTGAAATAAAGTATTTATCT
CCTGTCTACAGTTTAATAAATAGCATCTAGTACACATTCA

(R) 3' untranslated sequence beyond cDNA [SEQ ID NO. 19]

TTTTGTGTTG GATACTGTGT TAGGTGCTGG AGGAAAAAAG ATGAATAGAA
CATCTTCTAT GTACTTCATG CGCTCACAGT CTGGTTGTAG AGACTGTCAC
ATAAACATTT CATCCCAATT CATTTATTTG TTCATTCCCT CAGCCAATAT
ATATTGAGTT CTTACTCTGT GCCAAGAAGT GTACTACATT TCTGGGATTA
AGTGGATATA AGGAGATCTC AGTGTTTAAT CTGCCTGAGG GGAGACTAAA
TTAAGTGACA TGGAAACTTG GGTCTTGAAA AACATTTTAA GGTATTTTTT
TCTTTTCTCT CTCTCTCGCT CTGTCTTTCT CTCTCTTTCT TCAGGGTCTC
CCTCTGTTGC CCAGGCTGGA GTCAGTGGCA CTCATAGCTC ACTGCAGCCT
TGATCTCCTG GGCTCAAGAG TTCTTCCCAC CTCAGTCTCC TAAGTAGCTT
GGACTACGG

FIGURE 3K

-1108 GCTTGGCTC CAAAGGCTT GCGATTACAG GCGTGAACCA CTGGGCTAG CTTGTAGCA GCTTTTAAA TCGAGAGCA -1029
 -1028 TAAAGCTGTA TTTTGAAGG TTATGCAAG GATTCGAGT AGAATGAG TCAATTACAG ATCCCAATTA TTATCTTTC -949
 -948 TATTCGAGAA AGCTTTTTT TTTCTTTCG CACTCTGTT TATGAGAA ATGAGGTTT GCGTGGGTT TTAGGAGATC -869
 API Pa.1
 -868 AGCTAGATTC TTATGATCTG TCACATGCTT GGAATGTTGG GAGGATTTG GAGAGCTGA TGTACTTGT CTTAGATTCG -789
 -788 GATTTTAAAT TGAAGACAGAT GATCTTTAG GCGCATGCTA CTAATCTGAG GTTTTAGCA CAGAGTCACTA TCGAGAGCA -709
 -708 TCGAGACTTA GCGCATTTGT TCAATGCTG TCATAAATTA CCAGGCTGC TGTCTTGGT TACTTTTAA GACAGTTTCA -629
 -628 TCTGAGCTTT CTGGGATAT CTTCTTTGAG CAAAGCTCA TTAGGCTGG AATCTATTG TGTCTGAGG TATGACATAC -549
 AP3-RV SPI/R-AP3-1
 -548 TTGAGACCA CAGCTTATA AATATACAA GAATTTGGA GCTATGCTT TCTATGAT TATCTTTTA TAGCTTTTA -469
 -468 TGTTCATTAC GCGATGCTAG AGGTGAGAT CAGACACAA TATGAAATA GTTTCTAAG TGTGAGGAT AATCTTAC -389
 -388 AGGAGAGGG TCGAGAGAA TAAATGCCC CATATTTAA ATAGATAT TCGAGAGAA CAGTGGAGA GATGAGGCT -309
 PEA3 PEA3
 -308 GTGCTCTCC TTTTACTTC CAAAGAGAA AGTATGCTC TATTTAGTC ATAGATAT GAGGTCTT TATTGCTAC -228
 PEA3-RV API-RV API
 -227 CAAAGCAGT CTGTCAGAG TAAATAGCTA GTATTTGCTT AGCTGCTAC AGTCTGCA TACACAGAA AACTGCA -149

FIGURE 3L

-148 ATGCACTGCC TCTTCCTGC CTGCTACCG TCTCTCTCT CAGCAATCT ATCCCGCT CTCTCTCTA CCCAAATTT -69
 FEAL-RV SFI-RV

-68 CGAGCCCATC ACTGAGCTG ACTTCGCA TCTGATGA ATAAATCTAG-CACCCCTAT GGTGTGCA CACTTCTCTG +12
 H-APP-1 Ets-1RV AT-Rich Motif Exon 1 →

Figure 4

TABLE II
Exon-Intron Junctions of the Human Cathepsin-K Gene.

No.	CDNA (bp)	Donor	Acceptor	Intron size (bp)	Amino Acid interrupted
1	48	GCAGGtaacggttttgcaact....ctacattcttctcagcgtg		1437	Noncoding
2	169	CAAGgtgcctgggtctctg....ctatgctttgttttagctgg		462	Lys40/Val41
3	292	CATggcaagtatagtcttca....tctctttgttcttagACCA		85	Met81/Thr82
4	448	TCAGgtactctctcttctt....ctgtatccctttcagcgtc		1624	Gln133/Gly134
5	567	ACAGgtgagtgagattgct....gtttcttccagccagGAAG		4326	Gln206/Glu207
6	813	AAAGgtaaagaagctgctga....tagttgttatctctagctgt		270	Gly262
7	939	ACAGgtaatgaggggaaca....tgattggtctttacagCTGG		2270	Ser297

Exon and intron sequences are designated by upper and lower case letters respectively.

FIGURE 5

```

1
HumcatK .....HW GLKVL LLPV SFA.LYPEEI LOTHWELWKK THHKQYNHKV
RabOC-2 .....HW GLKVL LLPV SFA.LHPEEI LOTHWELWKK TYSKQYNHKV
HumcatS .....HKA LVCVLLVCS AVAQHKDPT LQHHHHLWKK TYGKQYKEKH
HumcatL .....HNPTL ILAALCLGIA S.ATLTPDHS LBAQHTKWKKA HHHHLY.GHH
HumcatH MWATLPLLCA GAWLLGVVPC GAAELSVNSL EKFIIFKSWMS KHHHTY:ST..
HumcatU ..... ..... .....HWQLNNS LCCILVIAHA
HumcatD .....MQP SSLPLALCL LAAPASALVR IPLHKFTSIR RTHHREVINHSV
HumcatE .....MKT LLLLLLVLE LGEAQGSLHR VPLRRHPSLK KKLHARSQ.I.
HumcatG .....MQP LLLLLAFLLP TGAEEGEI.. .....THHRE

51
HumcatK DEISRRL.IW EKNLKYISIH NLEASLGVHT YELAHNHLGD HTSEKVVQKH
RabOC-2 DEISRRL.IW EKNLKHISIH NLEASLGVHT YELAHNHLGD HTSEKVVQKH
HumcatS EEAVERRL.IW EKNLKFVHLH NLEHSHOHMS YDLOHNNHLD HTSEKVVQKH
HumcatL EECWRRRA.VW EKNHKKHIELH NQEYRECKHS FTHAHNAFGD HTSEKVFQVH
HumcatH EEEYHRLQTF ASNWRKINAH N....NGNHT FXHALNQFSD MSPAFIKNEY
HumcatS SRSPSEHFPVS DELVNYVNAH HTTNQAGHNF YNVDMSTLKR LCGTFI.....
HumcatD EDLIAKGPVS KYSQAVPAVT EGPIPEVLKN Y.HDAQYYGE IGIGTPPQGF
HumcatE SEFWKSHNLD MIQFTESCSH DQSAKEPLIN Y.LDMXYFGT ISIGSPPPQGF
HumcatG SRPHSRPYHA YLQIQSPAGQ SRCG.....G F.LVREDFVL TAAHCWGSNI

101
HumcatK TGLKVPLSHS RSNDTLYPE WEGRAP.DSV DYRKKG.YVT PVKHQGCQS
RabOC-2 TGLKVPPSR SRS NSNDTLYPD WEGRTP.DSI DYRKKG.YVT PVKHQGCQS
HumcatS SSLRVP.SQH QRNIT.YKSN PNRILP.DSV DWREKG.CVT EVKYQGCSCA
HumcatL NOFQ...NRK PRKGKVFQEP LFYEAP.RSV DWREKG.YVT PVKHQGCQS
HumcatH L.HSEPOHNS ATKSNYLRGT ..GPTP.PSV DWKXKGFVS PVKHQGCQS
HumcatB .....ODPK PPQRVHPTED LKLPASFDAR EQWPCPTIK EIRDDQSCNS
HumcatD TVVFDTCSSN LWVPSIHCKL LDIACWIIHK YNSDKS..ST YVKNHTSFH
HumcatE TVIFDTCSSN LWVPSVYCT. .SPACKTHSR FQPSQS..ST YSQPGQSFH
HumcatG NVTLG..... .....AHNIQRR ENTQQH..IT ARRAIR..HP

151
HumcatK CWAFFSSVGL EGQLKKTKR LLN..LSPQN LVDCVSE... NO..UCNGGY
RabOC-2 CWAFFSSVGL EGQLKKTKR LLN..LSPQN LVDCVSE... NY..CCGGGY
HumcatS CWAFFSAVGL EAQLKKTKR LVS..LSAQN LVDCSTERYO NR..GHHHGP
HumcatL CWAFFSATGL EQQHFRKTR LIS..LSEQN LVDC.SCPQC NB..GHHHGL
HumcatH CWTFTSTIGAL ESAIAIATCR HLS..LAZQQ LVDC.AQDFN NY..UCNGGL
HumcatB CWAFFGAVEAI SORICIHNA RVSVEVSAD LITCCGSHCG D...GHHHGY
HumcatD HYGSGSLSGY LSQDTVSVPF QSASSASALG GVKVERQVFC BATHKQITF
HumcatE QYGTGSLSGI ICADQVSV... .....E CLTVVQGPFC BATHKQITF
HumcatG QYNQRTIQND IHLQLSRR. ....VRRNRNVNP VALPHRQKH

201
HumcatK HTNAFQYVQK NRGIDSEDA ..... .....PYVQGES
RabOC-2 HTNAFQIVQR NRGIDSEDA ..... .....PYVQGE
HumcatS HTTAFFQYIID NRGIDSDASY ..... .....PYKANDL
HumcatL HDTAFQYVQD NGGLDSEESY ..... .....PYEATEE
HumcatH PSQAFEXILI NKGINCEDTY ..... .....PYQGEED
HumcatB PAEAWNE.WT RKGLVSGGLY ESHVGCNPYS IPPCEHNVNG SHPPCTCKH
HumcatD IAAKFDGIL. .GHAFFRIS VNNVLPVFDN LMQQKLVDQN IFSFY.SHH
HumcatE VDAEFDGIL. .GLOYPFLA VGGVTPVFDN HHAQNLVDLP HPSVVMSEH
HumcatG RPETLCTVA. .G..WGRVS KRRGTDTLRS VQLRVQRDRQ CLRIPG:YHP

```

FIGURE 5A

```

251
HumcatK SCH.....YNPTGKAAK CRGYREIPEG N.EKALKRAV ARVGHVSVVAI 100
RabOC-2 SCH.....YNPTGKAAK CRGYREIPEG N.EKALKRAV ARVGHVSVVAI
HumcatS KCQ.....YDSKYRAAT CSKYTELPGY R.EDVLKEAV ANKGIVSVGV
HumcatL SCK.....YNPKYSVAN DTGFVDIPK. Q.EKALKRAV ATVGHVSVVAI
HumcatH YCK.....FQPGKAICF VKDVANITTY D.EEAHVEAV ALYHVPVSPAF
HumcatB TPKCSKICEP GYSPTYKQDK HXCYNSTSVS NSEKDIHAEI YKHHPVVKIAP
HumcatD DAQPGGELML GGTDSKYKKG SLSTLVNTRK AYWQVHLDOV EVASGHVTLCK
HumcatE ZGGAGSELIF GGYDHSHEFG SLNWVPVTXQ AYWQIALDNI QVGHVTVHFGS
HumcatG RRQ.....ICVGR RERKAAFK..GDSGHVHJH

```

```

301
HumcatK DASLTSFQFY SKGVYIDESC ..NSDNLNHA VLAVGYGIQ. ...KGNKIMI 150
RabOC-2 DASLTSFQFY SKGVYIDENC ..SSDNVNH VLAVGYGIQ. ...KGNKIMI
HumcatS DARHPSFFLY RSGVYIEPSC ...TONVNHG VLVVGYGDL. ...HCKRYWL
HumcatL DAGHESFLFY KEGIYFEPDC ..SSEDHHDG VLVVGYGFES TESDHNKYWL
HumcatH EVTQD.FMHY RTGIYSSTSC HKTPDKVNHA VLAVGYG...EKHHTITWI
HumcatB SV.YSDFLLY KSGVIQHYTC EMHGG...HA IRELGHGVE. ...HGTIYWI
HumcatD EGCEA...IV DTGTSLVHGP VDEVNELQKA IGAVPLIQGE YHIFCEKYST
HumcatE EGCEA...IV DTGTSLVHGP VDEVNELQKA IGAVPLIQGE YHIFCEKYST
HumcatG NVARG...IV SYGKSSGVPP ....EVFTRV SSFLPWIRTT MR....SPKI

```

```

351
HumcatK IK.....NS WGENWGNKGY ILHARNKNNH CGIAN..LAS FPKH..... 400
RabOC-2 IK.....NS WGENWGNKGY ILHARNKNNH CGIAN..LAS FPKH.....
HumcatS VK.....NS WGHNFGEEGY IRHARNKNNH CGIAS..FPS YPEI.....
HumcatL VK.....NS WGEZNGHGGY VKHAKDRNH CGIAS..AAS YPTV.....
HumcatH VK.....NS WOPQHGHNQY FLIERGK.NH CGLAA..CAS YPIHIV.....
HumcatB VA.....NS HNTDWDGNGF FKILRGQ.DH CGIESEVVAG IPRTHQYWEK
HumcatD LPAITLRLGG KCYKLSPEDY TLKVSQAGKT LCLSGPHGMD IPRTHQYWEK
HumcatE HPDVTFTING VPYTLSPTAY TLDPFVQCHQ FCSSGFQGLD IPRTHQYWEK
HumcatG LDQMETPL.....

```

```

401
HumcatK .....428
RabOC-2 .....
HumcatS .....
HumcatL .....
HumcatH .....
HumcatB I.....
HumcatD LGDVFIGRYY TVFDRDNNRV GFAEAAAL
HumcatE LGDVFIGRYY TVFDRDNNRV GLAPAVP.
HumcatG .....

```

FIGURE 6

[SEQ ID NO: 20]

```

M W G L K V L L L P V V S F A L Y P E E
I L D T H W E L W K K T H R K Q Y N N K
V D E I S R R L I W E K N L K Y I S I H
N L E A S L G V H T Y E L A M N H L G D
M T S E E V V Q K M T G L K V P L S H S
R S N D T L Y I P E W E G R A P D S V D
Y R K K G Y V T P V K N Q G Q C G S C W
A F S S V G A L E G Q L K K K T G K L L
N L S P Q N L V D C V S E N D G C G G G
Y M T N A F Q Y V Q K N R G I D S E D A
Y P Y V G Q E E S C M Y N P T G K A A K
C R G Y R E I P E G N E K A L K R A V A
R V G P V S V A I D A S L T S F Q F Y S
K G V Y Y D E S C N S D N L N H A V L A
V G Y G I Q K G N K H W I I K N S W G E
N W G N K G Y I L M A R N K N N A C G I
A N L A S F P K M

```

Figure 7

Intron(s) Amplified	Oligonucleotide Sequence (5' to 3')	cDNA Location
1	CCGAAACGAAAGCCAGACAAC (S) GCTCACCACAGGTAGCAGCAG (AS)	exon 1 exon 2
2	CTGCTGCTACCTGTGGTGAGC (S) CCCAAATTAAACGCCGAGAG (AS)	exon 2 exon 3
3	CTCTCGGGCGTTTAATTGGG (S) GGTACTTTGAGTCCAGTCATC (AS)	exon 3 exon 4
4	CCAGACCTCTGTCGACTATCG (S) CACATATGGGTAGGCATCTTC (AS)	exon 4 exon 5
5 & 6	GAAGATGCCCTACCCATATGIG (S) GTTACATTATCGCTATTGCAC (AS)	exon 5 exon 7
7	GCAAAGGTGTGTATTATGATG (S) GCCGTTGTTCTTATTTCGAGC (AS)	exon 7 exon 8

SEQ ID NO.:21
SEQ ID NO.:22

SEQ ID NO.:23
SEQ ID NO.:24

SEQ ID NO.:25
SEQ ID NO.:26

SEQ ID NO.:27
SEQ ID NO.:28

SEQ ID NO.:29
SEQ ID NO.:30

SEQ ID NO.:31
SEQ ID NO.:32

ABSTRACT OF THE DISCLOSURE

The invention relates to cathepsin K polypeptides, polynucleotides encoding the polypeptides, methods for producing the polypeptides, in particular by expressing the polynucleotides, and agonists and antagonists of the polypeptides. The invention further relates to methods for utilizing such polynucleotides, polypeptides, agonists and antagonists for applications, which relate, in part, to research, diagnostic and clinical arts.